SAME WORDS PERFORMED SPOKEN AND SUNG: AN ACOUSTIC COMPARISON

Jaan Ross
Department of Arts, University of Tartu, Estonia
Estonian Academy of Music, Tallinn, Estonia

ABSTRACT

In a recent study (Ross and Lehiste 2001), we have described ways how, in singing, the prosodic structure of a language is modified and, vice versa, how the musical structure accommodates itself to requirements of the language. This paper compares the prosodic realization of phonetically identical words in speech and in singing. There are words in Estonian folksongs which are frequently repeated within a single performance as well as across different performances. These words function like a refrain. In some cases they represent a non-word like ['jumpʻ.per'jɑ:ː], in other cases a semantically meaningful word like ['nuǩ.ku] or ['nukʻ.ku] (‘doll’, gen. or part. sing. in Estonian). The acoustical analysis demonstrates that a refrain word in singing tends to retain a similar prosodic shape even across different performers while, at the same time, this shape may be fundamentally different from its normative pattern for the spoken language. This finding is remarkable because prosodic features (mostly duration) is used in Estonian in order to convey lexical and grammatical differences to listeners. It may be concluded that in singing, the prosodic structure of a language can in some cases be strongly subordinated to the musical structure, even though the result is unacceptable from the linguistic point of view.

1. BACKGROUND

The topic of our research is the relationship between the use of duration in the prosodic structure of the language, the metrical structure of the folksongs, and the realization of the quantity patterns in singing. We use the methodology of acoustic phonetics, namely spectrographic analysis of recorded samples of speech and song. We restrict the analysis to samples from a specific historical singing tradition characteristic of the majority of the so-called Baltic-Finnic cultures, including Estonian and Finnish. Peoples representing these cultures inhabit the eastern shores of the Baltic Sea in the north-eastern part of Europe. Their singing tradition is usually referred to as the Kalevala or runic singing tradition; sometimes these songs are also referred to as old folksongs, to distinguish them from newer folksongs that exhibit a different structure. A comprehensive ethnomusicological overview of the tradition is provided by Rüütel (2001).

From a typological perspective, Estonian and Finnish (as well as other Baltic-Finnic minority languages) are remarkable due to their unusually complex prosodic systems (Engstrand and Krull 1994). Both languages make lexical as well as grammatical use of contrastive duration Ð the feature of quantity. In Finnish, vowels can be contrastively short or long in any syllable, and consonants can be contrastively short or long (i.e. single versus geminate) anywhere except in word-initial and word-final position. The Estonian quantity system is even more complicated. In the first place, the system is hierarchical: duration is contrastive at the level of segments, syllables, and higher-level units, displaying a hierarchy of many-to-one mappings (different segmental combinations mapping into a smaller number of contrastive syllable durations, and contrastive syllable combinations mapping into higher-level units like prosodic feet). The second peculiarity of the Estonian system is the fact that it is ternary: there are three-way oppositions at every level Ð segments, syllables, and metric feet. The distinctive quantities are called short, long, and overlong, and are conventionally referred to as Q1, Q2, and Q3. Vowel duration is contrastive in the first syllable of a word. Consonants can be in the contrastive three quantities within a prosodic foot, and in a two-way opposition between prosodic feet and in final position; there is no durational contrast in consonants in word-initial position.

It is evident that in singing, the durational or quantity constraints inherent in the language have to compete with constraints from other domains than speech. These additional temporal constraints are, firstly, metrical, i.e. related to regular repetition of strong and weak timing units, and, secondly, rhythmical, i.e. related to more fine-grained and individual timing patterns characteristic to music (of which singing is a part). Thus the topic of our study is the temporal structure of Estonian folksongs looked at from three aspects: the prosodic structure of the language, the metric structure of the verse, and the rhythmic and melodic structure of the songs. In particular, we are investigating the ways in which these three aspects interact.

As to Estonian prosody, it has been a matter of debate at which level (phoneme, syllable, or prosodic foot or disyllabic sequence) the ternary quantity oppositions primarily operate. According to Lehiste (1997), the optimal level for the description of quantity oppositions is the prosodic foot. Among other characteristics, the foot as an entity has a tendency toward isochrony, which is achieved through complementarity in the duration of the two syllables of which it consists: short, long, and overlong first syllables are followed by successively shorter second syllables. The three contrastive quantity degrees are characterized by specific ratios between the two syllables (the S1/S2 ratio). For words in the short quantity (Q1), this ratio is approximately 2/3, for words in the long quantity (Q2), it is 3/2, and for words in the overlong quantity (Q3), the S1/S2 ratio is approximately 2/1. When the consonants are in short quantity, the S1/S2 ratio can be represented by the ratio between the vowels (the V1/V2 ratio). Note that in words of Q1, i.e. words with a short initial syllable, the unstressed second syllable is phonetically longer than the stressed first syllable. Unstressed second syllables can be open or closed; closed second syllables are considered to be long.

ISBN 3-931852-67-9
ISSN 1617-6847

2. AIMS

The aim of this study is to compare acoustical realization of single words in speech and in singing. In order to make possible to treat the data statistically, such words were selected for the study which are frequently repeated in sung productions. Those words function mostly as refrain words: they are placed at the end (rarely also at the beginning or middle) of the structural unit in folksongs, the verse line. Identical refrain words may also be frequently found in productions by different performers. Examples of the refrain words in Estonian folksongs are ['nuk̆.ku] or ['nukˑ.ku] ('doll', gen. or part. sing. in Estonian), or ['jumpˑ.per'jɑ:ː], a nonsense word.

A descriptive notation of a folksong line is presented below. The line consists of two parts: the first part is a regular eight-position verse line followed by the second part, the refrain word. The line originates from the recording of K.K., with the catalogue reference number, RKM Mgn II 394j, in the Estonian Folklore Archives in Tartu.

In the Folklore Archives, it is possible to obtain many field recordings where a sung performance of a folksong is preceded by its spoken counterpart. It means that the words of the song have been recited (and recorded) first and only thereafter the song is performed and recorded in full, including its melody. This situation is due to natural hesitation of the informant to sing in front of a microphone, in order be recorded. An informant may claim at first that she has forgotten the tune and is able to remember the words only, and only later to agree to sing. From the point of view of the present study, availability of parallel recited and sung texts by the same performer, which are recorded on the same day, is very favorable because it makes possible to compare spoken and sung productions recorded under maximally similar conditions.

We want to interpret phonetic differences between spoken and sung realizations of the same words: (1) how the emergence of 'distortions' of the words in singing (as compared to speech) may objectively be explained; (2) how do those 'distortions' project onto the normative prosodic system of the language; (3) how the 'distortions' may influence perception of those words by the listeners.

3. METHOD

Three recorded items were selected for analysis from the Estonian Folklore Archives in Tartu. Their catalogue reference numbers were ERA Pl 95 B1, RKM Mgn II 397f, RKM Mgn II 589b, RKM Mgn II 2049c. Items 1 and 2 were performed by M.S., item 3 by M.M. and item 4 by G.J., respectively. Item 1 was recorded in a studio on a shellac disk; the other three were recorded under fieldwork conditions using a magnetic tape recorder. All singers lived in the same South Estonian region, namely Karksi (parish). Copies of the original recordings were obtained from the Folklore Archives (with permission). The recordings of M.S. and M.M. contained sung performances only, while the recording of G.J. included, in addition, the recited version of the text, which was later recorded in sung form.

Copies of the original recordings were subjected to acoustic analysis. Durations of the individual phonemes were determined using parallel wide- and narrow-band spectrographic representations of the sound signal. A Kay Elemetrics Computerized Speech Laboratory (Model 4300) connected to a PC-computer was used for measurements with a sampling frequency of 10 kHz.

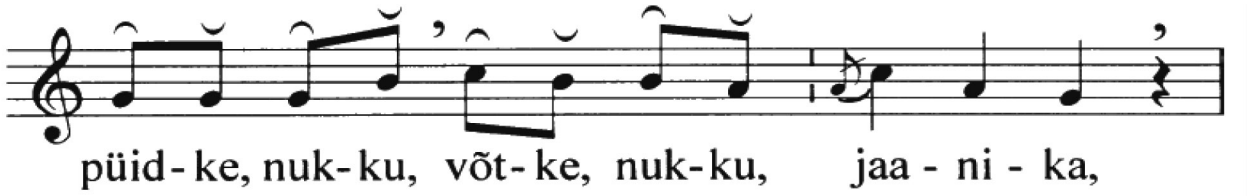

Figure 1: An example of an eight-syllable folksong line (the first bar) followed by a refrain word, [%j%%.nik%.k%] (the second bar).

4. RESULTS

4.1. Acoustical measurement of segment durations

The measurement results for three words, ['nuk̆.ku] (or ['nukˑ.ku]), ['jumpˑ.per'jɑ:ː] and ['seː.ri'jɑ:ː], are presented in Tables 1, 2 and 3, respectively. For the first word, it was possible to obtain data on its pronunciation comparatively in spoken and in sung productions of one performer, G.J., as well as in sung productions of the other performers, M.S. and M.M. For the second and third words, only the sung productions of the performer M.S. were available.

For the first word it was impossible to decide whether the performers intended to produce it as a Q2- or a Q3-word. It is a refrain word which is not linked syntactically to any other words in the text. If a Q2-word, it would be a genitive form of the nominative *nukk* ('doll'); if a Q3-word, it would be a partitive or an illative form of the same nominative. The second and third words have no definitive meaning in the language. Note also differences in tempo between the three performers. M.S. has sung considerably slower than M.M. and G.J.

ISBN 3-931852-67-9
ISSN 1617-6847

4.2. Normative expectations from speech prosody

Current phonetic theory (Lehiste 1997) describes differences in spoken short, long and overlong disyllabic words in terms of the ratio of the duration of the first syllable to the duration of the second syllable. This S1/S2 ratio is expected to be approximately 2/3 for the Q1-words, 3/2 for the Q2-words, and 2/1 for the Q3 words. Table 1 shows that the average ratio for the spoken production of the word ['nuǩ.ku] by G.J. was 1.17, which suggests that the word is likely a Q2 word, not a Q3 word Đ in other words, the genitive form, not a partitive or an accusative form. The value of 1.17, however, is even smaller than the expected 1.5, but individual differences of this magnitude have been described earlier and are common in spoken language (Ross and Lehiste 1994).

The S1/S2 values for the same word in sung production fall within the range between .63 to .70, i.e. they exhibit the pattern close to a Q1-word. This is clearly incompatible with the phonological structure of the word which by no means fits the Q1 pattern. We have to conclude that the normative phonetic pattern of the word ['nuǩ.ku] is disregarded in sung performances by all the three performers. It is remarkable that the numerical S1/S2 values in singing are close to each other for all the singers.

It is a bit more complicated to discuss the words ['jumpˑper'jɑːː] and ['seːˑri'jɑːː] because these words do not exist in the language. Intuitively, both words might be expected to be compound words consisting of a disyllabic word ['jumpˑper] or ['seːˑri] followed by a monosyllable ['jɑːː]. If this treatment is correct then the first disyllable ['jumpˑper] would be a Q3-word and the second disyllable ['seːˑri] a Q2-word. The S1/S2 value (1.07) for the Q3-word in singing (see Table 2, the rightmost column) is much below the expected value of 2.0. The S1/S2 value (1.16) in singing (see Table 3, the rightmost column), however, is not incompatible with spoken Q2-words (cf. Table 1, last row, the rightmost column).

As to the monosyllabic word ['jɑːː], it is phonologically considered to be a Q3-word. This means that the vowel in it is overlong also in spoken production and, indeed, even more so in singing (see Tables 2 and 3, penultimate column). It should be mentioned that such a word is ideally suited for completing a structural unit both in speech and in singing, because of the so-called final lengthening. This phenomenon has been described earlier as a signal to the listener of the presence of a boundary between two structurally meaningful units (Sundberg and Verrillo1980).

	Performer	N	[n]	[u]	[k]	[u]	S1	S2	S1/S2
SUNG	G.J.	10	99 (9.3)	113 (14.2)	169 (25.1)	387 (78.5)	297	472	.63
	M.S.	8	112 (14.5)	289 (48.7)	435 (28.2)	754 (100.9)	628	972	.65
	M.M.	24	77 (10.2)	90 (12.2)	195 (22.6)	282 (67.1)	265	380	.70
SPOKEN	G.J.	8	64 (15.0)	83 (14.5)	204 (34.9)	110 (30.7)	249	212	1.17

Table 1: Comparison of sung and spoken realizations of a single word, [‘nuǩ.ku] or [‘nukˑ.ku] (‘doll’, gen. or part. sing. in Estonian). The sung realizations are produced by three performers and the spoken version by a single performer. Acoustical durations are given for each phoneme in milliseconds (standard deviation in parentheses). Syllable durations are also presented (so that the geminate [k] is evenly divided between the two syllables. The S1/S2 ratio is in the rightmost column. N is the number of items measured.

N	[j]	[u]	[m]	[p]	[e]	[r]	[j]	[%]	S1	S2	S3	S1/S2
11	87 (11.8)	157 (44.9)	124 (24.3)	243 (34.4)	199 (18.9)	216 (39.5)	102 (17.9)	775 (89.6)	530	496	877	1.07

Table 2: The sung realization of a single refrain word ['jumpˑper'jɑːː] by the performer M.S. This is a non- word. Acoustical durations are given for each phoneme in milliseconds (standard deviation in parentheses). Syllable durations are also presented. Since the first component of this compound word is overlong, the geminate [p] is divided between S1 and S2 so that 2/3 of it belongs to S1 and 1/3 to S2. The S1/S2 ratio is in the rightmost column. N is the number of items measured.

N	[s]	[e]	[r]	[i]	[j]	[%]	S1	S2	S3	S1/S2
11	136 (15.9)	377 (24.0)	74 (22.7)	386 (24.7)	108 (91.3)	809 (171.0)	513	460	917	1.16

Table 3: The sung realization of a single refrain word ['seːˑri'jɑːː] by the performer M.S. This is a non- word. Acoustical durations are given for each phoneme in milliseconds (standard deviation in parentheses). Syllable durations are also presented. The S1/S2 ratio is in the rightmost column. N is the number of items measured.

ISBN 3-931852-67-9
ISSN 1617-6847

4.3. Descriptive notation of the rhythm of words

It is essential that the folksong tradition we investigate is an oral one. However, for all those songs the excerpts from which we have analyzed above, there exist descriptive notations which have been compiled by ethnomusicologists. We next discuss the rhythmic patterns which have been used in order to describe the timing in the three refrain words.

The word ['nuǩ.ku] is always situated at the end of a line (do not confuse this situation with the verse line in Figure 1 which is not representative in this respect). In rough terms, it is always performed as consisting of a short note following a longer note in singing, and consequently notated as well. The normative rhythmic patterns in notation depend on many factors including, e.g. the meter. This way, the short-long rhythmic succession may be denoted in the score by an eighth value following a quarter (the productions of G.J. and M.M.), or a quarter following another quarter with a fermata (the production of M.S.)

Notation of the rhythm in words ['jumpˑ.perˈjɑːː] and ['seˑ.riˈjɑːː] in the production by M.S. is more uniform in comparison with the first lexical example. The meter for the notation of these productions is usually chosen as 4/4, and the two words are accompanied by a rhythmic pattern consisting of two quarters plus a half note.

5. CONCLUSIONS

The following conclusions can be made from the results of the present study.

- There is a tendency towards all the notes (syllables) in the melody being isochronous. This tendency is evident with disyllabic words like ['jumpˑ.per] and ['seː.ri] which in sung realizations obtain the average S1/S2 values of 1.07 and 1.16. As a Q3- and a Q2-word, respectively, their corresponding S1/S2 values should be close to 2.0 (Q3) and 1.5 (Q2).

- Transformation of the word ['nuǩ.ku] in singing departs significantly from the pattern of its spoken counterpart. Its S1/S2 value of 1.17 in speech is replaced by the value of .63 to .70 in singing. This means that, in phonetic terms, a Q2-word is changed into a Q1-word. Although such a Q1-word with a short plosive [k] would have no match in the spoken language (['nu.ku] is a non-word in Estonian), the switch of a phonetic pattern for Q2 to the one of Q1 is remarkable. It demonstrates that disyllabic sequences in the folksong performance may not only strive to isochrony but, in some cases, may obtain a shape which is openly contradictive from the speech-prosodic point of view.

- There is a clear tendency in the analyzed songs the final syllable (note) in a line to be lengthened. In the case of a monosyllabic word [%j%%%] the outcome is fully acceptable for the speech prosody (the vowel in the monosyllabic word is overlong). In the case of the second syllable of a disyllabic word ['nuǩ.ku] falling in the line-final position, the outcome is unacceptable from a linguistic point of view.

The above conclusions are in agreement with our earlier study of the temporal structure of Estonian folksongs (Ross and Lehiste 2001) and in certain aspects (a Q2-word obtaining a Q1-word pattern in singing) deepen our understanding about the co-variative behavior of the speech prosody and the musical rhythm in singing.

To what extent those results may be applied to other languages than the Estonian? There is relatively little research on comparative aspects of speech prosody and musical rhythm. It is well known that in Chinese language(s), different F0 patterns are employed in order to convey lexical and grammatical differences, somewhat analogically to how the duration differences are explored in the Baltic-Finnish languages. It seems as the relationship between the different tone patterns in Chinese language and their combination with the melodic structures in singing is no less complex than the interaction of prosodic and rhythmical duration patterns in the Estonian folk songs (Stock 1999).

6. REFERENCES

1. Engstrand, O. and Krull, D. (1994). Durational correlates of quantity in Swedish, Finnish and Estonian: Cross-language evidence for a theory of adaptive dispersion. *Journal of Phonetics 51, 80-91*.

2. Lehiste, I. (1997). Search for phonetic correlates in Estonian prosody. In I. Lehiste and J. Ross (eds), *Estonian Prosody: Papers from a Symposium* (pp. 11-35). Institute of Estonian Language: Tallinn.

3. Ross, J. and Lehiste I. (1994). Lost prosodic oppositions: A study of contrastive duration in Estonian funeral laments, *Language and Speech 37, 407-424*.

4. Ross, J. and Lehiste, I. (2001). *The Temporal Structure of the Estonian Folk Songs*. Mouton de Gruyter: Berlin and New York.

5. Rüütel I. (2001). Estonia, traditional music. In: S. Sadie (ed), *Grove Dictionary of Music and Musicians*, 2nd edition (pp. 342-347).

6. Stock, J. (1999). A reassessment of the relationship between text, speech tone, melody, and aria structure in Beijing Opera. *Journal of Musicological Research 18, 183-206*.

7. Sundberg, J. and Verrillo, V. (1980). On the anatomy of the ritard. *Journal of the Acoustical Society of America 68, 772-779*.

ISBN 3-931852-67-9
ISSN 1617-6847