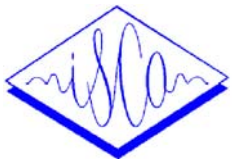

Speech Prosody 2004
Nara, Japan
March 23-26, 2004

Prosody in Speech and Singing

Ilse Lehiste

Department of Linguistics
The Ohio State University
ilsele@ling.ohio-state.edu

Abstract

The study deals with the question of the manifestation of the prosodic system of a language (here Estonian) in singing. Previous research has shown that durational contrasts are neutralized in the sung version of the folksongs. The present study makes a first attempt to establish whether there is a similar neutralization of prosodic oppositions with regard to melody: whether the melody of the song interacts with contrastive pitch, or whether melody overrides prosodic contrasts based on pitch in the spoken form of the language.

1. Introduction

The topic of this paper is the possible interaction between speech prosody and melody in sung Estonian folksongs. The study is an extension of a series of previous studies, carried out in collaboration with a musicologist (Dr. Jaan Ross). A summary of our research was published recently in book form (Ross and Lehiste 2001). In that study, we concentrated on the relationship between contrastive *durational patterns* in speech, and musical *rhythm*. The current paper takes a first look at the possible relationship between *melody* and speech prosody.

2. Prosodic structure of the language

The studied material is taken from one of the Estonian folksongs analyzed in the course of the previous study. In order to explicate how the melody aspect might affect the realization of contrastive prosodic patterns, a brief survey of that previous study will be offered first.

We studied the interaction of three prosodic systems: the prosodic system of the language, the metrical structure of the folksongs, and the musical structure of the oral presentation by folksingers. The language – Estonian – has a rather complicated prosodic structure involving both contrastive duration (quantity) and fundamental frequency. There are three degrees of quantity, called short, long, and overlong, or Q1, Q2 and Q3. Vowels have three contrastive durations in the stressed first syllable of a word. Consonants have three contrastive durations between a stressed and an unstressed syllable. The domain of quantity patterns is a disyllabic sequence or foot. There is a kind of foot isochrony, with a characteristic relationship between the durations of the two: a short first syllable is followed by a half-long second syllable, and the duration of the second syllable decreases as that of the first syllable increases, so that the overlong first syllable is followed by the shortest second syllable. The basic relationships between the syllables are 2:3 for Q1, 3:2 for Q2, and 2:1 for Q3.

The second prosodic system that is involved in this interaction is the metrical structure of the folksongs. The metre is shared by Finnish and Estonian folksongs and is commonly referred to as the Kalevala metre. A line consists of four trochaeic metric feet, with metrical stress on the first syllable of each foot. Another way of describing the structure of the line is by referring to positions within the line: odd-numbered positions are stressed, and even-numbered positions are unstressed. There are two quantity rules that govern the placement of words into the line: a short stressed first syllable is excluded from a position with metrical stress, and a long or overlong first syllable *must* occur in a position with metrical stress. This results in a contradiction between word stress and metrical stress in the case of words beginning with short open syllables, since primary stress is always on the first syllable of a word (with the exception of some recent loanwords).

The third system that is involved in singing the folksongs is the musical rhythm. The line consists of eight notes; each note corresponds to one syllable. These notes are commonly transcribed as eighthnotes and are assumed to have equal duration, but in fact the notes corresponding to the syllables carrying metrical stress are significantly longer, thus manifesting musical rhythm.

A sung folksong represents the results of the interaction of these three systems. We showed in our studies that musical rhythm overrides the durational oppositions. In particular, differences in vowel durations are neutralized. This means that all vowels have more or less the same duration, regardless of whether they are stressed or unstressed in the spoken form of the language, or whether they are in what is referred to as Q1, Q2, or Q3. A contrast in intervocalic consonants remains, but it is reduced to a two-way short-long opposition: the difference between long and overlong consonants is likewise neutralized.

3. The role of fundamental frequency

During these earlier studies we looked only at duration and rhythm. The present study takes a look at the potential interaction between duration, rhythm, and melody. The long and overlong syllables in the analyzed language differ not only in duration, but also in the fundamental frequency pattern: in spoken disyllabic sequences, a long first syllable has a level or rising pitch contour, while an overlong syllable has a falling pitch contour. (I have treated this phenomenon as an indication that Estonian may be on the way to changing from a quantity language to an accent language, where pitch participates in the prosodic system; cf. Lehiste 2003).

The aspect we did not study was the question whether there is any interaction between the melody and the prosodic structure of the language. The question is more general and applies to all tone languages: what is primary, the melody of

10.21437/SpeechProsody.2004-2

the song or the contrastive tones of the text? Is the tonal structure of the text subordinated to the melody, or are words chosen so that, for example, high tones correspond to high notes in the melody? I am not talking about translating and adapting the text of a well-known opera aria, but about what happened at the stage when folksongs were created - usually by anonymous representatives of a language and culture.

As already mentioned, pitch plays a part in the prosodic structure of Estonian, contributing to the difference between long and overlong degrees of quantity. In the spoken parts of the interviews on the tapes which we analyzed, the singers maintained a distinction between long and overlong quantities both with respect to duration and with respect to the associated fundamental frequency contours. In spoken language, the fundamental frequency contour is level or rising on a contrastively long vowel, and falling on a contrastively overlong vowel. It appeared interesting to check whether a rising melody would favor the realization of the sung word as a word in the long quantity, and whether a falling melody might favor words in overlong quantity – to the extent that the neutralization of the two quantities might be less complete.

It has to be kept in mind that in Estonian folksongs, each syllable corresponds to a single note, and each note corresponds to a syllable, so that a disyllabic word is given two notes. The pitch difference between the two quantities, however, is realized within the first syllable. The second syllables of both word types are expected to be low. So the rising or falling melody would not reflect the pitch patterns directly, but might be suggestive. In particular, a falling melody might favor a more speech-like realization of words in overlong quantity, since in spoken language, overlength is associated with a falling pitch contour.

4. Materials

I selected a lengthy recorded song from the extensive set analyzed for the previous study (Ross and Lehiste 2001, pp. 154-158). The song consists of 108 lines, containing 432 metric feet. Table 1 offers average durations of syllable nuclei in words in the two quantities, as they were sung with a rising, level, or falling melody on the two notes. The ratio between the syllable nuclei of both word types is also given, as this constitutes one of the defining characteristics of the manifestation of contrastive word-level prosody. In spoken language, words in long quantity (Q2) would have a ratio of 3:2, and words in the overlong quantity (Q3) would have a ratio of 2:1.

Table 1. *Average durations, in msec, of V1 and V2 in long (Q2) and overlong (Q3) words sung to rising, level, and falling melody. The words had the structure (CVVCV) for Q2, and CVVVCV for Q3. Number of occurrences in parentheses.*

Melody		Q2		Q3
Rising	V1 (6)	344	V1 (9)	385
	V2	253	V2	269
	V1/V2	1.36		1.43
Level	V1 (4)	322	V1(14)	363
	V2	278		247
	V1/V2	1.16		1.47
Falling	V1 (6)	405	V1 (9)	414
	V2	264	V2	294
	V1/V2	1.53		1.41

5. Interpretation of results

If there is an association between the direction of the melody and the realization of the words, overlong (Q3) words should have a longer first syllable with falling melody than with rising melody. But there is very little difference in the average durations of overlong syllables, regardless of the direction of the Fo movement (the number of instances is too small to make a statistical reliability test meaningful). In fact both long and overlong words had their longest first syllables in the case of falling melody.

Another characteristic of spoken overlong words is the relative shortness of the second syllable. If the hypothesis holds and falling melody suggests overlength, the Q3 word should be realized with a shorter second syllable in the case of a falling pitch movement, but in fact the second syllables of Q3 words with falling pitch were longer than when Q3 words were sung to a melody with rising pitch. And the durational structures of rising Q3 words and falling Q2 words are practically identical.

As regards the ratios between syllable durations, no regularities can be observed in the present set of data. It should be kept in mind that all ratios, as well as durational differences between first and second syllables, would become smaller when the duration of the intervocalic consonant is added to the duration of V2. Since the musical notes in the folksong melody have approximately the same duration, the short vowels of the second syllable are always lengthened, and any significant differences between them are neutralized. This affects Q3 words in particular, since in spoken words of this quantity the overlong first syllable is twice as long as the short second syllable. None of the ratios were even close to the 2/1 ratio that is characteristic of words with overlong first syllables. Let us recall that the Q2 words have ratios of 3:2, or 1.5; in these materials, both Q2 and Q3 words had ratios close to 1.5 with both rising and falling melody. It might be significant that both Q2 and Q3 words had the smallest difference in V1 and V2 durations (and ratios closest to 1.0) when the melody was level – neither rising nor falling.

Thus the results of this exploratory study turned out to be negative. Melody appeared to have no influence on the realization of a contrast that is at least partly based on tonal

movement in the spoken form of the language. It should be remembered that the durational difference between long and overlong words was likewise neutralized in singing.

The role of fundamental frequency in Estonian is comparable to that of languages with so-called contour tones. It may well be that the interaction of contour tones with melody differs from a possible interaction of melody and tone in languages where each syllable has contrastive tone, and each syllable corresponds to a musical note. But it is equally possible that musical structure neutralizes tonal oppositions – in the same way as musical structure neutralizes durational oppositions in a language like Estonian.

Fundamental frequency and stress.

The role of fundamental frequency is not restricted to contrastive pitch patterns; among its functions is contributing to the identification of a word (or distinguishing between a word and a non-word). In the language under consideration, word stress is always on the first syllable, and higher pitch is one of the features associated with stress. Since this is a language that uses contrastive duration in its prosodic system, duration is not freely available for marking the presence of stress; thus it may be expected that other features – such as higher pitch – may be relatively more important in signalling the presence of stress.

The role of F0 is fairly straightforward: under normal conditions (i.e. when the sentence is not produced with an exceptional intonation contour) a stressed syllable is marked by higher fundamental frequency. Thus in disyllabic words, the first, stressed syllable has higher pitch than the second, unstressed syllable.

It appeared interesting to check whether the identificational role played by pitch in its association with stress was in any way discernible in the materials under consideration. Of course the melody determines the fundamental frequency of each syllable, but the creator of the folksong could have positioned the words in the line in such a way that higher notes coincided with stressed syllables.

Let us recall here that the quantity rules of the folksong place stressed long and overlong first syllables in positions of metrical stress (or ictus), and exclude stressed short first syllables from ictus position. If the F0 level is correlated with presence of word-level stress, one would expect long and overlong first syllables to occur in positions where the melody has a higher note than the notes associated with neighboring unstressed syllables. The question of words with stressed short first syllables is of particular interest, since the metrical structure requires these syllables to fall in metrically unstressed positions, but the role of F0 in identifying stressed syllables would require that these syllables, too, have higher pitch than neighboring syllables.

Table 2 presents the observed pitch relationships between word-initial syllables and immediately following syllables. Only word-initial metric feet are considered. The Q1 words (with a short stressed first syllable) included in this table actually fail to obey the quantity rules, but such deviations from the regular pattern are not unusual.

Table 2. *Pitch level on word-initial stressed syllables relative to following syllables. H-L = stressed syllable has higher pitch, L-H = stressed syllable has lower pitch, E-E = both syllables are sung on the same note.*

Word type	H – L	L – H	E – E	Total
Q1	0	21	18	39
Q2	33	19	22	74
Q3	43	28	37	108

As can be seen from the table, stressed long and overlong first syllables occurred on higher notes of the melody in 43% instances of the occurrences of Q2 and 40% of the occurrences of Q3. The behavior of Q1 words is anomalous insofar as in this particular song, there were no instances where a stressed short syllable, here produced with metrical stress, would have had higher pitch than the following syllable.

Of greater interest is the behavior of those stressed short syllables that were produced in conformity with the quantity rule: occurring in the non-ictus part of a metric foot, so that the following syllable carries the ictus of the next metric foot. Table 3 offers an overview of these cases.

Table 3. *Pitch on Q1 syllables in non-ictus position, relative to the syllables occurring in preceding and following ictus positions. H = higher pitch, L = lower pitch.*

Preceding Ictus		Q1 syllable		Following ictus	
H	L	H	L	H	L
15	4	5	14	3	16

As can be seen from the table, the short syllable that carried word-level stress, but lacked metrical ictus, was sung with a lower pitch than the preceding syllable bearing metrical stress. The following syllable with metrical stress had likewise lower pitch in the majority of instances.

These results indicate that word stress was not manifested through higher pitch as a possible stress correlate. In other words, the syllable that carries stress and is normally spoken at a higher pitch was not helped by fundamental frequency to retain its stressedness in singing.

Melody thus appeared to have no influence on the realization of a contrast that is at least partly based on tonal movement in the spoken form of the language. Pitch level did not contribute to the identification of the presence of word stress, as it does in speech. As was mentioned above, durational contrasts were likewise neutralized in vowels. In singing, melody and rhythm dominated over linguistic prosody.

6. Further outlook

The role of fundamental frequency in Estonian is comparable to that of languages with so-called contour tones. It may well be that the interaction of contour tones with melody differs from a possible interaction of melody and tone

in languages where each syllable bears contrastive tone, and each syllable corresponds to a musical note. But it is equally possible that musical structure neutralizes tonal oppositions in the same way as musical structure neutralizes durational oppositions in a language like Estonian.

As already mentioned, the results of this exploratory study turned out to be negative. Melody appeared to have no influence on the realization of a contrast that is at least partly based on tonal features in the spoken form of the language. It should be remembered that the durational difference between long and overlong words was likewise neutralized in singing.

The role of fundamental frequency in Estonian is comparable to that of languages with so-called contour tones. It may well be that the interaction of contour tones with melody differs from a possible interaction of melody and tone in languages where each syllable has contrastive tone, and each syllable corresponds to a musical note. But it is equally possible that musical structure neutralizes tonal oppositions – in the same way as musical structure neutralizes durational oppositions in a language like Estonian.

I am fully aware of the limitations of the reported study. I believe, however, that the question of a possible relationship between contrastive tone in a language and the melody of folksongs created in that language is worth a more detailed investigation.

Acknowledgement

I am grateful to Dr. Jaan Ross for insightful comments.

7. References

[1] Lehiste, Ilse , 2003. Prosodic change in progress: from quantity language to accent language. In: Paula Fikkert and Haike Jacobs (eds.), *Development of Prosodic Systems.* Berlin – New York: Mouton de Gruyter. 47-65.

[2] Ross, Jaan, and Ilse Lehiste, 2001. *The Temporal Structure of Estonian Runic Songs.* Berlin – New York: Mouton de Gruyter.