ARCHIVES OF ACOUSTICS
13, 3–4, 269–278 (1988)

PERCEPTION ASPECTS OF A RULE SYSTEM FOR CONVERTING MELODIES FROM MUSICAL NOTATION INTO SOUND

L. FRYDÉN

Conservatory of the Swedish Radio, Edsberg, Sweden

J. SUNDBERG, A. ASKENFELT

Department of Speech Communication and Music Acoustics Royal Inst. of Technology
(S–10044 Stockholm 70, Sweden)

The starting point of this project is the "mechanical" impression obtained when a computer rather than a good musician converts the musical notation into the corresponding sound sequences. A basic assumptions is that this mechanical effect can be partly eliminated by introducing "pronunciation rules", which introduce minute, context dependent deviations from the durations, pitches an amplitudes specified in the musical score. The method applied is analysis by synthesis. A computer program developed for the conversion of text to speech (Carlson and Granström [3]) is applied to the singing synthesizer MUSSE (Larsson [4]). In the present, musical version of the conversion program, the input is the melody in musical notation, and the output is the melody, synthetically performed. A set of such pronunciation rules have been formulated for, and tested on traditional Western melodies. One group of rules operates with short time windows of two or three notes. Thus they are applied using as criteria the size of the musical interval formed by adjacent notes, or the difference in duration between two or three adjacent notes. Another type of rules operates with a time window of variable length; the window is limited by the distance between adjacent chord changes. The rules manipulate the duration and amplitude of notes. In other words all these rules introduce discrepancies between what is written in the notation and what is actually played. The effects of these rules are evaluated by means of listening tests with musically sophisticated judges. The results show that the musical acceptability of such synthesized performances can be substantially improved by applying these "pronunciation" rules. Furthermore, our experiences indicate that the magnitude of these discrepancies between what is written and what is supposed to be played is very critical. If the magnitude is so high, that the effect is identified for what it actually is, then the effect generally tends to be musically impossible. If, on the other hand, the magnitude is too small, there is, of course, no effect at all. It seems that the musically useful effects are found in between this threshold of diagnosis and the threshold of perception. The perceptual implications of these rules regarding musical communication will be discussed, and the effects of the rules may be demonstrated by means of tape illustrations.

Praca przedstawia propozycję rozwiązania problemu „sztuczności" i „martwoty" charakteryzujących muzykę produkowaną w sposób syntetyczny przez komputery. Autorzy opierają się na swym długoletnim doświadczeniu w analizowaniu muzyki „żywej" i obserwacji wszelkich odchyleń jej parametrów w zakresie od matematycznie regulowanych równomiernych skal. Proponują wprowadzenie całego szeregu reguł rozbijających w określony sposób równomierność skal w zakresie czasu, natężenia i częstości. Podano opis eksperymentu psychoakustycznego, w którym grupa doświadczonych słuchaczy porównywała i oceniała kompozycje wyprodukowane syntetycznie z użyciem i bez użycia wspomnianych reguł. W większości przypadków wprowadzenie odchyleń od równomierności nadało muzyce bardziej żywy charakter i zostało pozytywnie ocenione przez słuchaczy. Dokonano analizy uzyskanych wyników i wyciągnięto wnioski co do dalszych ulepszeń proponowanego systemu.

1. Introduction

When music is being played, a string of note signs is converted into a sequence of tones. If this conversion is realized in a one-to-one fashion, e.g. by a computer, the result is musically a disaster. This fact raises a question what are the exact relationships between note signs and their corresponding acoustic signals in musical performances? Or, in other words, what do musicians do when they play, and why do they do that?

Previous studies of musical performance have revealed an almost overwhelming complexity (see e.g. BENGTSSON and GABRIELSSON [2]). SEASHORE (1938) measured the actual durations of the tones in several musical performances of the same song. Among other things he found that in none of the singers was there "a slightest approach to an even time for a measure". Also, he found that musicians tend to show a modest degree of reproducibility as regards the performance of a given piece of music.

These findings would seem trivial to a musician. He knows perfectly well that a piece of music can be played in a number of different ways which are all musically acceptable. And yet, the field of liberty for musicians is by no means unlimited; there is a large class of performances which are musically inacceptable. This suggests the existence of *musical performance rules* stating what is musically permitted and what is not in the conversion of notes to tones. If a player violates these rules, he runs the risk of being classified as a poor musician.

These performance rules are scientifically interesting. First of all, they separate music and non-music. Second, the rules would bear witness of auditory, perceptual and cognitive functions involved in the ears' and the brain's processing of the acoustic signal in music listening. Thus the rules may reveal some of the basic requirements for enjoying music.

2. Method

If we want to explore the rules which a musician must normally obey in playing a piece of music we can choose measurements or synthesis-by-rule. In the present project, the latter strategy was chosen. The reason for this is our assumption that the rules are numerous and that they interfere with each other. Also, the pedagogical and artistic experience of one of the authors (LF) had generated a number of hypotheses regarding such rules.

Like in a previous study of musical performance focussing on singing (SUNDBERG [7]) a computer-controlled vowel synthesizer was used (MUSSE, LARSSON [4]) which can generate one part only. In the present experiment the signal characteristics were adjusted so as to be similar to a wind instrument timbre. The pitch frequency changed in steps in accordance with the equally tempered scale, and there was a very slight vibrato only. The amplitude changed in steps of 1/4 dB. The duration was controlled in steps of a time unit corresponding to 0.8 to 1.2 csec depending on the tempo; according to VAN NOORDEN [5] this is accurate enough.

The computer programs used for controlling the synthesizer were 1) a notation program (Askenfelt [1]) by means of which the melody can be written in ordinary music notation on the computer screen; and 2) a text-to-speech program written by CARLSON and GRANSTRÖM [3]. The information encoded in the notation is translated into "vowel sounds" of corresponding duration, pitch frequency and amplitude. The rules are triggered by specific sequences of durations and/or pitches, and they manipulate amplitudes and durations.

3. Rules

Up to now we have formulated a dozen rules and tested seven of them in a listening experiment. We will now present these rules, one by one.

(1) *Pitch and amplitude*. In almost every music instrument, the amplitude increases somewhat with the fundamental frequency, as is illustrated schematically in Fig. 1. Our rule increases the amplitude as function of frequency by 4 dB/octave.

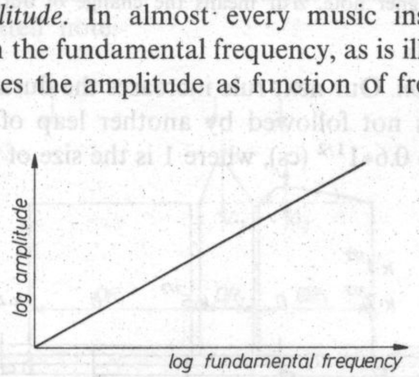

Fig. 1. Schematic illustration of the rule increasing the amplitude of the note as function of its fundamental frequency. The quantity is 4 dB of amplitude increase per one octave rise in fundamental frequency

(2) *Note value and duration.* This rule sharpens the duration contrasts between note values. All whole-notes and half-notes are played with their nominal durations while all quarternotes and eightnotes are played 1 cs too short, and every second sixteenthnote is played 1 cs too short. Thus long notes are played in their full length, and short notes are shortened a percentage which is inversely proportional to the nominal duration, (see Table 1). No compensation is made for the resulting perturbation of the mechanical meter.

Table 1. Shortening of the durations for different note values

Note value	Shortening csec	%
Sixteenth	−0.5	−3.1
Eight	−1.0	−3.1
Quarter	−1.0	−1.6
Half	0	0
Whole	0	0

(3) *Pitch increase and duration.* This rule simply states that the duration of each note is decreased by 1 csec the following note has higher pitch. In sequences of rising pitches the rule has the effect of raising the tempo somewhat, (see Fig. 2). As before, no compensation is made for the resulting perturbations of the meter.

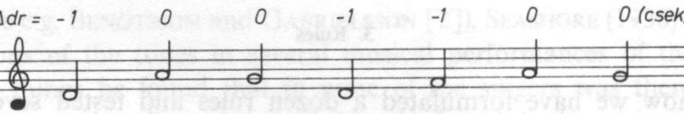

Fig. 2. Schematic illustration of the rule decreasing the duration of all notes which are followed by a higher note. Δdr means the change of duration

(4) *Leaps and duration.* Our next rule increases the duration of tones terminating a melodic leap which is not followed by another leap of identical direction. The quantity was adjusted to $0.6 * 1^{1/2}$ (cs), where 1 is the size of the interval in semitones, (see Fig. 3).

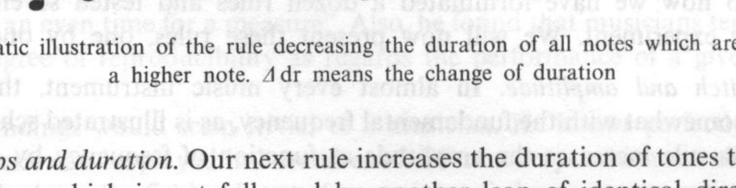

Fig. 3. Schematic illustration of the rule increasing the duration of all notes which terminate a melodic leap. 1 is the size of the leap interval in semitones

(5) *Leaps and pauses.* In instrumental music, particularly when played on bowed instruments, wide melodic leaps are often performed with a micropause during the pitch change. This is the effect of a rule, which decreases the amplitude of the final portion of the tone at a constant rate. The decrease starts at 0.5∗1 cs from the end of the note, where 1 is the interval size in semitones. This rule has a negligible effect on narrow intervals as major and minor second but is quite noticeable in wide intervals, (see Fig. 4).

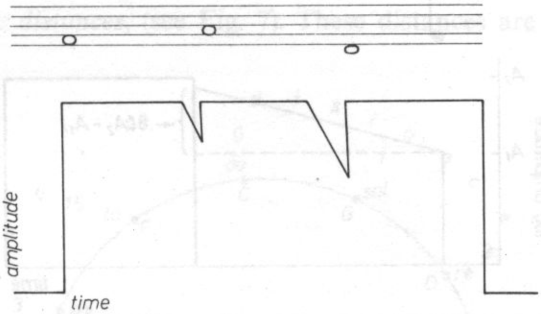

Fig. 4. Schematic illustration of the rule which inserts amplitude decreases towards the end of a note. The decrease has a constant rate. The onset time of the decrease is determined by the interval between the note and the following note

(6) *Note value contrasts and accents.* This rule marks contrasts in note values by accents, i.e. small and very rapid increase-decrease gestures in the amplitude. The rule adds an amplitude increment inversly proportional to the duration of the tone. The details of the accents are shown in Fig. 5. The rule adds an accent in two cases, as exemplified in the same figure. One is a short note surrounded by longer notes. The other case is a note terminating a specific pattern of changes in durations: a decrease followed by an increase. This rule has a clear effect particularly on the short notes after a dotted note.

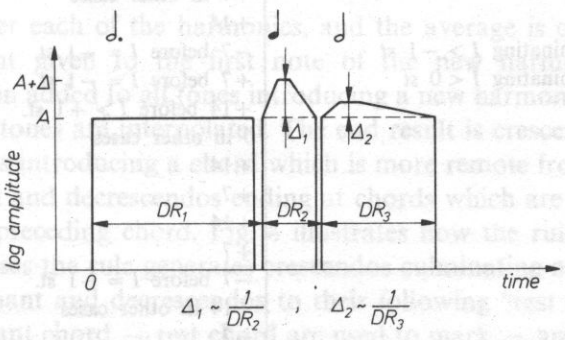

Fig. 5. Schematic illustration of the accents marking certain contrasts in note value

(7) *Amplitude continuity.* If the amplitude of tones changes stepwise, the melodic continuity may be disturbed. Our next rule states that the last amplitude reading of a tone should be corrected by a constant corresponding to 80% (in log. terms) of the amplitude difference between consecutive notes, (see Fig. 6).

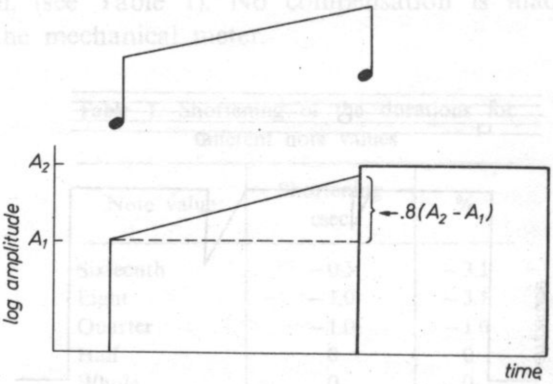

Fig. 6. Schematic illustration of the rule decreasing the amplitude contrasts between adjacent notes

(8) *Scale tone tuning.* The equally tempered scale is rarely used in music practice (SUNDBERG [9]). This rule modifies the scale tone frequencies by multiples of 7 cents which is our incremental step and which is close to one or two DL: s of pitch in a normal listener (RAKOWSKI [6]). Table 2 shows the values used.

Table 2. Deviations in cents from the equally tempered scale. I = interval, st = semitone

Scale tone in semitones above the root of the tonic	Deviation (cent)
1	-14 before $I = -1$ st $+7$ in other cases
2	$+14$
3 terminating $I > -1$ *st*	-7 before $I = -1$ st.
terminating $I < 0$ *st*	$+7$ before $I = -1$ st $+14$ before $I > +1$ st. 0 in other cases
4	$+14$
5	$+7$
6	$+14$
7	$+7$
8	-7 before $I = -1$ st. $+14$ in other cases
9	$+7$
10	-7
11	$+14$

The rules presented so far operate on a narrow time window of two or three notes only. The following rules use a wider time perspective, namely changes of harmony. The first note appearing over a new harmony is marked in the score. The following rules use such marks cues.

(9) *Chord changes and amplitude.* This rule generates crescendos and decrescendos. The technicalities are as follows. The distance along the circle of fifths from the root of the tonic is first determined for each note. Then, this distance is multiplied by 1.5 for scale tones located on the subdominant side of the circle. We have called the resulting values *tonic distances*, (see Fig. 7). These distances are averaged over all

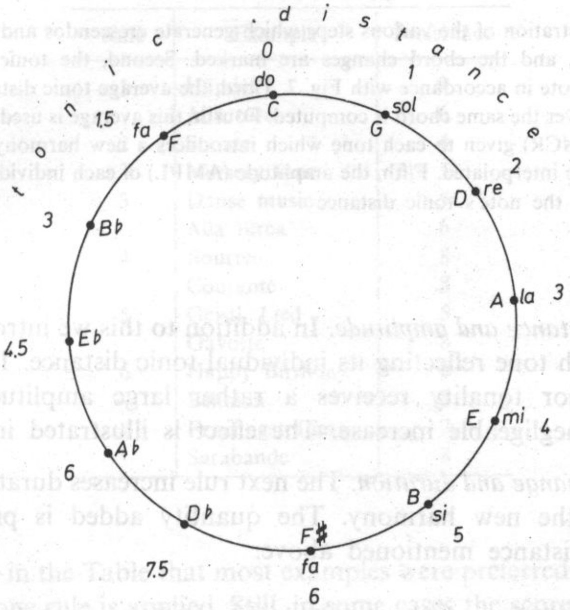

Fig. 7. Definition of tonic distance by means of the circle of fifths

tones appearing over each of the harmonies, and the average is converted into an amplitude increment given to the first note of the new harmony. When such increments have been added to all tones introducing a new harmony, the amplitudes of the intermediate tones are interpolated. The end result is crescendos culminating at harmonic changes introducing a chord which is more remote from the tonic than the preceding chord and decrescendos ending at chords which are less remote from the tonic than the preceding chord. Fig. 8 illustrates how the rule treats a typical cadence. In most cases the rule generates crescendos culminating at chords with the function of a dominant and decrescendos to their following "rest chords". As such sequences of dominant chord — rest chord are used to mark — among other things — the termination of phrases and subphrases in simple tunes, this rule often has the effect of marking the melody structure (cf SUNDBERG and LINDBLOM [8]).

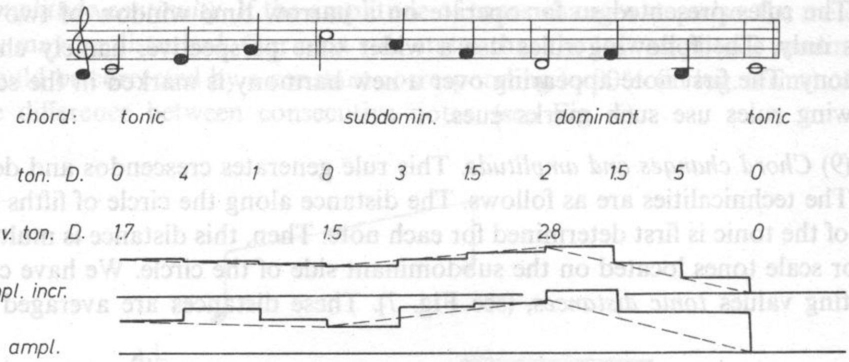

Fig. 8. Schematic illustration of the various steps which generate crescendos and diminuendos. First, the chords are identified, and the chord changes are marked. Second, the tonic distance (TON. D) is determined for each note in accordance with Fig. 7. Third, the average tonic distance (AVTON. D.) over the notes appearing over the same chord is computed. Fourth, this average is used to decide the amplitude increment (AMPL. INCR) given to each tone which introduces a new harmony: the amplitudes of the intermediate tones are interpolated. Fifth, the amplitude (AMPL) of each individual note is increased by an amount reflecting the note's tonic distance

(10) *Tonic distance and amplitude.* In addition to this we introduce an amplitude increment on each tone reflecting its individual tonic distance. This means that e.g. *F #* in *C* major tonality receives a rather large amplitude increase, while a G receives a negligeable increase. The effect is illustrated in the same Fig. 8.

(11) *Chord change and duration.* The next rule increases duration of the first note appearing over the new harmony. The quantity added is proportional to the averaged tonic distance mentioned above.

(12) *Phrase endings and duration.* This rule applies to and therefore mark phrase endings. As yet, we have no algorithm, detecting phrase boundaries automatically, so we mark the boundaries between phrases and subphrases in the input score. Our rule simply adds 4 csec to the note terminating a phrase. For notes terminating subphrases the last 1 cs is used for a pause.

4. Evaluation

A listening experiment test was carried out in order to test the effect on musical acceptability of each of the rules specified above. Thirteen melodies were presented in pairs of performances. In one performance no rule was applied i.e. the performance involved no changes of the nominal durations and amplitudes, while in the other performance one of the seven rules applied. The melodies where chosen so as to expose particularly clearly the effect of the rule to be tested. The tape was presented over headphones to 9 musically highly experienced judges who were asked to mark on a paper which performance in each pair they preferred from a musical point of view. The results are shown in Table 3.

Table 3. Result of listening experiment, where 9 musical experts compared the musical quality of synthetic performances. The numbers indicate how many of the experts who preferred the rule controlled performance. The examples are specified in the **Appendix**

Rule	Example	Preference
1	Händel	9
	Alla turca	6
2	Sonate	4
	Nursery tune	5
3	Danse music	5
	Alla turca	6
4	Bourrée	8
	Courante	8
5	Geistl. Lied	5
	Gavotte	8
6	Happy Birthday	5
	Bellman	8
7	Fruelingstraum	7
	Sarabande	8

It can be seen in the Table that most examples were preferred by more than half of the jury, when one rule is applied. Still, in some cases the scores are quite low. In some cases this low score probably had an unexpected explanation. When rule number 5, which introduces pauses during leaps, was tuned loudspeakers were used, while headphones were used in the test. It seemed that the absence of room reverberation in the headphone listening made the pauses sound too long.

There is a trend that excerpts from the baroque era are rated higher than other excerpts. It cannot be excluded that the reason for this is that our rules are perhaps optimized with respect to this kind of music. However, we believe that the reason is that our examples from the postbaroque periods are more dependent on the harmonic progressions underlying the melody; such examples are likely to suffer more from the fact that none of these seven rules reacts to the harmonic context.

It seems fair to conclude from the test that in a majority of cases each of the seven rules improve the musical acceptability of the performance. Still, a single rule does not appear to improve the quality of a performance to any appreciable extent. Rather the effect is that the mechanical impression of the performance is eliminated.

It is likely that the effect of the individual rule is dependent on the effect of other rules. Therefore we would expect that the summed effect of all rules will improve the performance much more than a single rule.

An important question is to what extent the exact formulation of our rules is critical; can the deviations from nominal durations and the changes in amplitude be replaced by a corresponding randomization? An informal listening test was made where two performances were compared. The first one was in accordance with the rules, while in the second one the durations and amplitudes were varied randomly within the limits set by our rule system; a random number between 0 and 1 was used to determine the quantity of the effects. The result demonstrated that the last mentioned, randomized performance appeared to lack stability in some way. This is not astonishing. It seems obvious that no event, which may catch a devoted listener's attention in an ideal performance, can be explained as random. It appears axiomatic that all which can be noticed in an ideal performance serves the purpose of communication.

5. Discussion

If it is correct to assume that our rules all contribute to the performance of a melody, an interesting question is what purpose these rules serve. Some of the rules probably reflect shear convention while some other might be dependent on properties of human perception and cognition. In any case it is interesting to speculate a little about the possible background of some of the rules.

We believe that some rules derive from the human voice. The reduction of amplitude steps at tone boundaries and the general growth of amplitude with fundamental frequency may be examples of this. Other rules are likely to have a psycho-acoustic background. The principle of lengthening the note which terminates a leap might reflect certain effects studied by Van Noorden [5]; the disruption of a melodic line caused by a wide leap can be eliminated by reducing the speed of playing. A similar effect might be obtained if the tone terminating the leap is lengthened.

There may also be purely psychological foundations for some of the rules. The lengthening both of the notes terminating a leap and of the first note appearing over a new harmony may serve the purpose of convincing the listener that the note, even though unexpected, was indeed intended. Also, the increase in tempo during sequences of rising intervals may have a psychological background. A rising interval may possess an "activating" connotation and is often associated with an upward locomotion. Then, the player may need to stress that this "upward motion" is not hard to execute, and he announces this by increasing the tempo somewhat. Another possible explanation would be that pitch rises are often combined with an increases tempo in excited speech. The rule lengthening the last tone in a phrase would be recognized by speech researchers; in speech this phenomenon is well known and is generally referred to as the principle of final lengthening. Our acquaintance with speech may have programmed us so that lengthening is interpreted as a sign of a final element. If so, the reversion of the rule (shortening of final note in a phrase) would hardly be possible in any performance.

There are, no doubt, more rules than those presented above, particularly if we turn to the performance of polyphonic music. Thus, we do not pretend that our rules are sufficient to produce a musically acceptable performance. Neither do we pretend that the present formulation of the rules is definite and different repertoires will probably need at least partly different rule systems. What we pretend is that the musical quality of a performance can be significantly improved by applying a set of rules.

Another important point is to state that we have not tried to model the multitude of choices which is available to the living musician and which allows him to play the same piece in many different ways, all of which are equally acceptable from a musical point of view. There are, on the other hand, some possibilities to include such a liberty. One way is to allow for personal variations in quantity of the rules. Another possibility is to vary the structural interpretation of the music which the musician performs. Also, we would like to declare that we do not believe that our performance rules must always be obeyed in a good performance. On the contrary, such a performance may be boring in the long run. We believe that musicians should violate one or more of the performance rules as soon as they want to tell something in particular or excite the audience by means of a surprise.

In our work with formulating rules we have observed that the quantity of the rule's effect is very critical. If the effect is too great, the effect becomes obvious in a physical sense, and then the effect is embarrassing. For example if the lengthening of notes terminating leaps is too large, one can hear that this is what happens, and the result is musically unacceptable. The musically proper effect of the rule arises when its effect is just noticeable but not indentifiable for its nature. Perhaps this is something which is essential for art in general: we do not want to be disturbed by information about the technical means behind the piece of art, we just want to enjoy it!

We repeat that these attempts to interpret the rules represent pure speculation. The point is that some of our rules probably have a background of some kind which may be independent of music. We believe that further research on this background will be interesting and rewarding.

6. Conclusions

From the above we conclude the following:

1. It is possible to improve the musical acceptability of a performance by applying a limited set of "pronounciation" rules.

2. Such rules can be discovered by means of a analysis-synthesis approach.

3. Such an approach enables us to formulate new hypotheses as to how the rules tried should be complemented and thus to contribute to knowledge about and scientific understanding of music.

Appendix

Origin of the melody excerpts used in the evaluation test:

J. S. BACH: *Gavotte* from Partita E major for violin solo, BWV 1006; *Courante* from Suite D major for cello solo, BWV 1012; *Bourreé* from Suite C Major for cello solo, BWV 1009; *Sarabande* from Suite c minor for cello solo, BWV 452; *"Dir, dir, Jehova ..." Geistl. Lied* BWV 452.

C. M. BELLMAN: *Vila vid denna källa*, Nr 82 in Fredmans epistalar.

G. F. HÄNDEL: *Sonata E major* for violin and continuo, op. 1, no. 15.

K. JULARBO: *Livet i Finnskoga* (*Dance music*).

W. A. MOZART: *Alla turca* from Sonate für Klavier A major, K 331.

F. SCHUBERT: *Frülingstraum* from Winterreise, op. 89, no. 11, D 911; *Sonate* für Violin und Klavier, op. posth. 137:1, D 384.

Traditional: *Happy birthday*.

References

[1] A. ASKENFELT, *Automatic notation of played music: the VISA project*, Fontes Artis Musicae, **26**, 109–120 (1979).

[2] I. BENGTSSON, A. GABRIELSSON, *Rhytm research in Uppsala*, In: Music room acoustics, Publ. 17 issued by the Royal Swedish Academy of Music, Stockholm, pp 19–56.

[3] R. CARLSON, B. GRANSTRÖM, *A phonetically oriented programming language for rule description of speech*, In: Speech Communication, [ed.] G. Fant, Almqvist and Wicksell, Stockholm, vol II, pp 245–253.

[4] B. LARSSON, *Music and Singing Synthesis Equipment MUSSE*, Speech Transmission Laboratory Quarterly Progress and Status Report 1, 38–40 (1977).

[5] L. P. A. S. VAN NOORDEN, *Temporal coherence in the perception of tone sequences*, Druk vam Voorschooten (Diss).

[6] A. RAKOWSKI, *Pitch discrimination at the threshold of hearing*, Proceedings of the 7th International Congress on Acoustics, **3**, 373–376 (1971).

[7] J. SUNDBERG, *Synthesis of singing*, Swedish Journal of Musicology, **60**, 1, 107–112 (1978).

[8] J. SUNDBERG, B. LINDBLOM, *Generative theories in language and music descriptions*, Cognition **4**, 99–122, (1976).

[9] J. SUNDBERG, *In tune or not? A study of fundamental frequency in music practice*, Speech Transmission Laboratory, Quarterly Status and Progress Report 1, 49–78 (1982).

Received September 13, 1988.