Three applications of analysis-by-synthesis in music science

Johan Sundberg

[1] Department of Speech, Music and Hearing
School of Computer Science and Communication
KTH Department of Linguistics, Stockholm University
[2] University College of Music Education Stockholm

Abstract. The article describes how my research has applied the analysis-by-synthesis strategy to (1) the composition of melodies in the style of nursery tunes, (2) music performance and (3) singing. The descriptions are formulated as generative grammars, which consist of a set of ordered, context-dependent rules capable of producing sound examples. These examples readily reveal observable weaknesses in the descriptions, the origins of which can be traced in the rule system and eliminated. The grammar describing the compositional style of nursery tunes composed by A. Tegnér demonstrates the paramount relevance of a hierarchical structure. Principles underlying the transformation from a music score file to a synthesized performance are derived from recommendations by a violinist and music performance coach, and can thus be regarded as a description of his professional skills as musician and pedagogue. Also in this case the grammar demonstrates the relevance of a hierarchical structure in terms of grouping, and reflects the role of expectation in music listening. The rule system describing singing voice synthesis specifies acoustic characteristics of performance details. The descriptions are complemented by sound examples illustrating the effects of identified compositional and performance rules in the genres analysed.

1 Introduction

In our time it is easy to collect data about physical properties of sounds and obtain a complete specification of their acoustic properties. Basically, just a microphone, a computer and some (freeware) analysis software is needed.

By contrast, the human hearing is quite selective. First, psychoacoustic masking can make acoustic properties of a sound inaudible, thus making these properties totally irrelevant to the sound perceived (see e.g., Gauffin & Sundberg, 1974). Second, our hearing system typically discards sound properties that do not convey useful information; for example, Japanese-speaking people typically fail to hear a difference between the consonants /l/ and /r/, which are acoustically quite different, but not used as different phonemes in the Japanese language.

This gap between the acoustic description of a sound and how it is perceived can be a problem. If one wants to describe the sound of an operatically trained voice, the description needs to include all properties that are perceptually relevant. In that situation, analysis-by-synthesis is a wonderful tool. After deriving regularities from relevant parameters, these regularities are used as the input to the synthesis. The synthesis sounds as the analysed voices only if all perceptually relevant parameters were included in the input. Thus, you plug your data into a synthesiser, listen to how it sounds and check if the synthesis sounds like the original.

Analysis-by-synthesis is a classical method in scientific research in general. The strategy is illustrated in terms of a block diagram in Figure 1. The starting point is reality (right) with characteristics that generate data. The scientist formulates hypotheses about the regularities that can be assumed to underlie these data and converts them to a model that is capable of predicting data. The scientist then compares these predictions with the observed data. As long as there is a difference between prediction and data, the hypotheses are modified accordingly, used for making new predictions, and then another tour can be made in the feedback loop. When eventually the only discrepancy between data and prediction is stationary noise, knowledge has been generated.

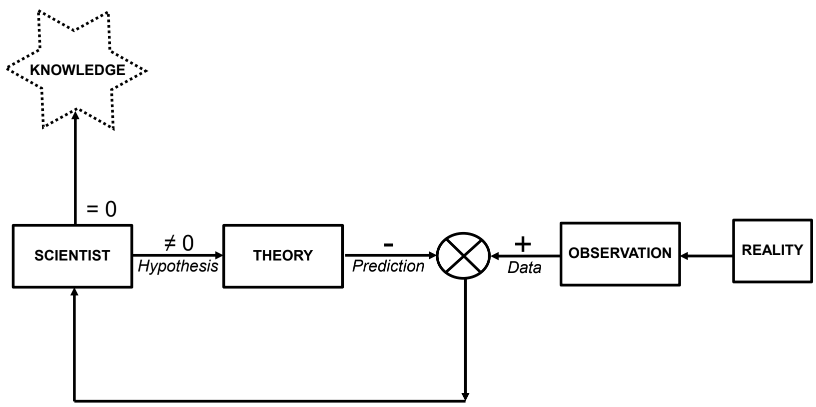

Fig. 1. Block diagram illustrating procedure of scientific inquiry. (After Figure 1 in Sundberg & Lindblom, 1976)

In the following, I describe how I have used the analysis-by-synthesis strategy in three specific applications: composition of melodies in the style of nursery tunes (Sundberg & Lindblom, 1976), music performance (Sundberg, Askenfelt, & Frydén, 1983; Friberg, 1995), and singing (Sundberg, 2006).

2 Compositional style

For the analysis-by-synthesis of the compositional style of nursery tunes I had the privilege of a joyful cooperation with Björn Lindblom, professor of phonetics and jazz pianist, but at that time doctoral student at Gunnar Fant's department at the Royal Institute of Technology KTH, in Stockholm, Sweden. He had just spent time at The Massachusetts Institute of Technology (MIT), where he met Ken Stevens and Morris Halle. This brought him into contact with the idea of generative grammar of phonology, a precisely formulated set of ordered rules that can generate as output all sentences and only sentences of a language. When he returned to KTH he was interested in applying this idea to music. The target was to formulate a grammar that could generate nursery tunes.

Figure 2 illustrates how the analysis-by-synthesis strategy was applied. The initial grammar rules were derived from musical analysis, complemented by assumptions derived from intuition. An example of the latter was a syntax that produces a constituent structure, which, in turn, is the input to prominence rules. These rules produce a prominence contour, which controls timing rules generating durations of the sequence of notes. The prominence contour also controls the chord rules that produce chord sequences. The prominence contour, the note durations and the chord sequences serve as the compound input to the tonal rules, which generate harmonised melodies.

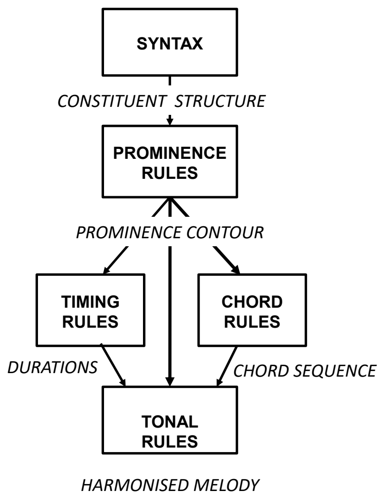

Fig. 2. Block diagram of the generative grammar of A. Tegnér's nursery tunes. (After Figure 6 in Sundberg & Lindblom, 1976)

Next, this process will be illustrated in detail in terms of an attempt to derive an existing nursery tune. Then, some melodies generated by random choices between alternatives offered by each rule in the grammar will be presented.

The songs we selected for analysis belong to a set of nursery tunes composed by the Swedish composer Alice Tegnér and published between 1892 to 1934 (Tegnér, 1892). Of this material we selected all 8-bar tunes in 4/4 time.

The first step was to assemble observations of regularities in terms of what events did occur and which did not occur in the different positions along the 8-bar period. We found that half notes and quarter notes followed by a pause *did* occur in the final stressed beats of bars 2, 4, 6 and 8, while eighth notes and dotted eight notes *did not* occur in these positions in these same bars (see Figure 3).

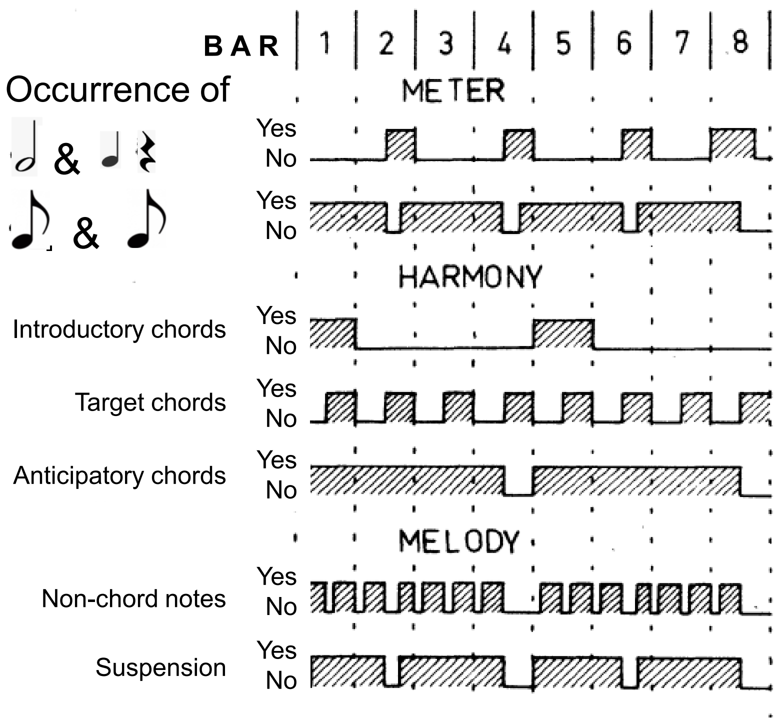

Fig. 3. Examples of regularities in terms of occurrencies in the positions of the 8 bar periods. (After Figure 5 in Sundberg & Lindblom, 1976)

Regarding chords, we found introductory chords first in the phrase and subphrase, while target chords (chord preceded by its dominant with or without its minor septime) occurred at the end of each bar in the 8-bar period. Anticipatory chords (dominant chords followed by its own tonic) occurred in these bars but not in the end of bars 4 and 8. With regard to melody, non-chord notes were not observed in stressed positions in the bars (beats 1 and 3). Also suspension notes showed a regular pattern.

Thus, we could see a number of regularities with respect to occurrence of various types of note values, chords and melody tones. The regularities were now organised into a generative grammar with ordered rules.

2.1 Test 1

As a first step we asked the question: "Is the grammar capable of generating a melody that Alice Tegnér actually composed?" For this test we selected the tune *Dansa min docka*. Its constituent structure was perfectly symmetrical and hierarchical, from the 8-bar period at the top, down to verse foot and, at the bottom, the four beats in each bar (Figure 4).

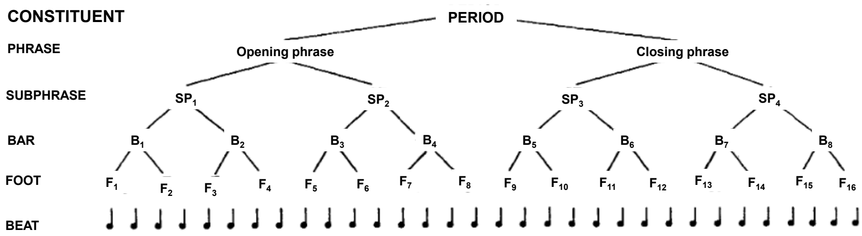

Fig. 4. Constituent structure of the 8 bar nursery tunes. (After Figure 7 in Sundberg & Lindblom, 1976)

As illustrated in Figure 5, the constituent structure was first converted into a prominence contour embracing the entire series of crotchet note positions in the melody. Each of the constituents shown in Figure 4 was then surrounded by a pair of parentheses, from the crotchet notes and down to the entire 8-bar period. After this, the innermost parenthesis pairs were erased at each hierarchical level, whereby one note was prioritised, keeping its prominence value, while the prominence of the remaining notes in the parenthesis pair fell by one step. The right-most note was prioritised at all levels except at the lowest one, where the left-most was prioritised. The bottom line in Figure 5 shows the result of the procedure. This prominence contour served as the entrance ticket to the generation of note values, chord sequences and melody.

Figure 6 illustrates the procedure for converting the prominence contour into a sequence of note values. As mentioned, the target to be generated by the rule system was a specific tune by Tegnér. The first step was to decide its scheme of sequencing; in this tune, as well as the other of Tegnér's 8-bar tunes we analysed, the metrics in the first 4-bar sub-phrase was copied in the second 4-bar sub-phrase. The next rule, the *catalexis* rule, imported from the theory of

Fig. 5. Derivation of prominence values by applying the rule *Erase innermost parentheses and prioritize left or rightmost prominence value.* (After Figure 8 in Sundberg & Lindblom, 1976)

verse, implied elimination of the last 4th note in the melody. The subsequent rules allowed dividing 4th notes into two 8th notes in certain positions and also dotting of 4th notes (a possibility not used in the target melody). After these steps, the metric pattern of the first sub-phrase was finished and then copied to the second sub-phrase. Last, these prominence values were translated into note values.

Fig. 6. Derivation of note values by applying the indicated metric rules. (After Figure 9 in Sundberg & Lindblom, 1976)

The prominence contour was the input also to the process that was used for derivation of chord progressions, as illustrated in Figure 7. First, notes with prominence value lower than 4 were deleted, such that each bar had only one prominence value left. In the following steps, the pattern of chord functions was determined (introductory, resting, anticipatory and target chords), their distance from the tonic (how many chord changes Tegnér used for a chord before arriving back to the tonic) and finally the chords themselves.

The derivation of a harmonised melody used all the above as its input: the prominence contour, the note values, and the chord sequence. The procedure is illustrated in Figure 8. After copying the sequencing pattern, pitches were determined for prominences 1, 2, and 3. For the remaining prominences, suspensions were allowed (not used in the target melody) and two principles applied: (1) *harmonic implication* determined if the tone should belong to the prevailing chord or not, and (2) *adjacency principle* states that a tone must be given a

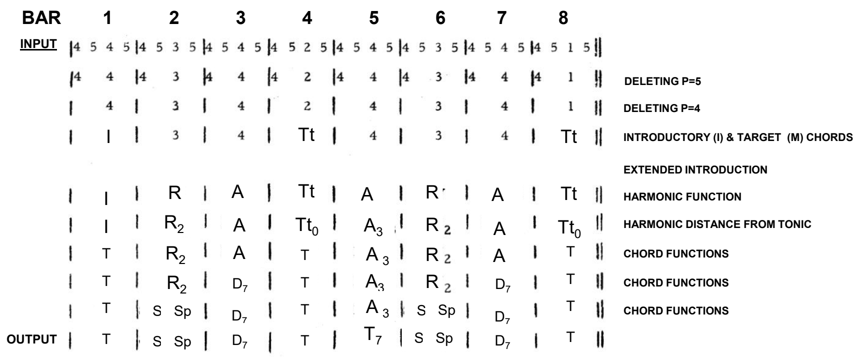

Fig. 7. Derivation of chord progressions; I=Introductory chord, Tt=target chord, R=resting chord, A=anticipatory chord; indices represent the harmonic distance to the tonic; chords: S=subdominant, Sp=parallel of subdominant, D7=dominant seventh, T7=tonic seventh, T=tonic. (After Figure 11 in Sundberg & Lindblom, 1976)

pitch lying between its two already decided neighbours. Finally, the sequencing pattern was applied.

The bottom line in the figure shows the resulting melody (stems down) together with the original Tegnér tune (stems up). Differences occurred in few places but did not sound as violating the compositional style.

The result of Test I thus showed that our proposed grammar passed the test rather well; it generated the major properties of a melody composed by Tegnér. Obviously, it is possible to analyse the deviations between the original and the grammar's version and to use them for perfecting the grammar, i.e., for another tour in the feedback loop shown in Figure 1. We did not do this, but instead proceeded to perform the second test of the grammar.

2.2 Test 2

The second test was to find an answer to the question, "Does the grammar generate only tunes in the compositional style of Tegnér's nursery tunes?" Here, the method was to use the rule system in the same way as described above, but this time to make a random choice among the alternatives allowed by each of the rules. The system passes this test if the generated nursery tunes sound as if Tegnér had composed them.

Figure 9 shows some tunes generated in this way. We asked jazz pianist and composer Bengt Hallberg to make and play arrangements of them (Sound example 1). Being a good friend of Björn, he answered Björn's question about what he thought about the melodies "Well, I have been writing arrangements of many poor melodies in my life".

Our working with the grammar offered us some quite clear examples of what was causing a problem. For instance, before adding the rule *adjacency principle*,

Fig. 8. Stepwise derivation of a harmonised melody from the input top line. In the output melody (bottom line), downward stems show the original version, and upward stems the version generated by the grammar.(After Figure 13 in Sundberg & Lindblom, 1976)

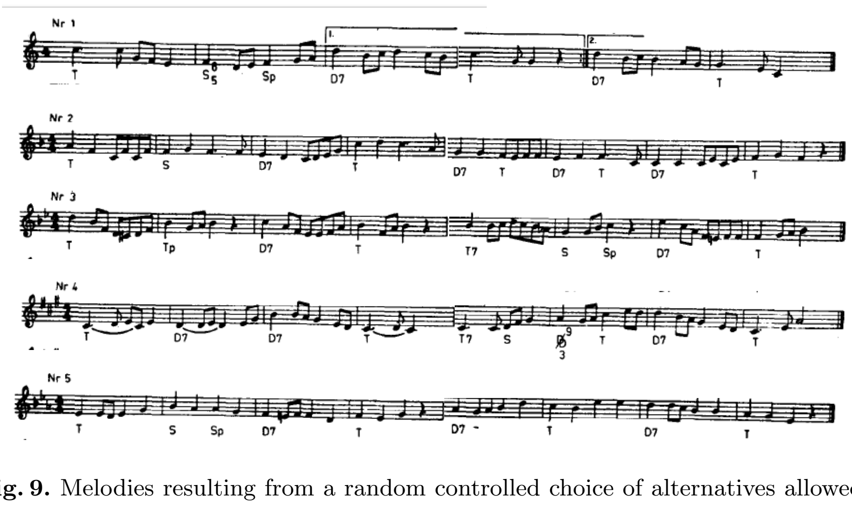

Fig. 9. Melodies resulting from a random controlled choice of alternatives allowed by the rule system

some melodies contained several big pitch jumps. Here, it was easy to hear what the problem was and to solve it by adding that rule. Another problem, less easy to solve, was the frequent occurrence of the pitch F4 in melody 2, and the odd-sounding ending of melody 5. Problems of these types offered hints on how to further develop the description of the characteristics of the compositional style of Tegnér's nursery tunes by another turn in the feedback loop in Figure 1, a possibility that we did not use.

We did not undertake a formal test to find out if these melodies sound as nursery tunes composed by Tegnér. However, we noted that most people who heard them found that they sound like nursery tunes and that they were easy to remember.

3 Music Performance

It is well-known that the performance is an exceedingly important property of music. The deviations between the score´s nominal specifications of duration, pitch and dynamics have been analysed since the days of Seashore (1967) and found to be quite extensive. Since the 1970s, research on music performance has been a productive branch of music science. It has revealed important and meaningful regularities regarding musicians' meaningful deviations from the score (see e.g., Gabrielsson, 2003).

A sequence of random factors brought me to research in this area. In 1976, a student at KTH, Björn Larsson, asked me if he could build an electric organ as his thesis work. This was a question that I already had heard from several other students, but given my interest in the voice as a music instrument in singing, I managed to talk him into building a singing synthesiser. The resulting machine, the Music and Singing Synthesis Equipment (MUSSE) was finished a year later (Larsson, 1977). It was played from a keyboard, and with properly adjusted formant frequencies and bandwidths and with a sinusoidal vibrato, it sounded quite realistic.

A year later, another student appeared, whom we convinced to build an interface between MUSSE and the main lab computer CD 1700. This combination caught the interest of my department colleagues and friends Rolf Carlson and Björn Granström, who had developed a rule-based text-to-speech synthesis system, RULSYS (Carlson & Granström, 1975, 1976). It consisted of ordered, context-dependent rules that generated speech from a text input. They constructed a musical version of this system for score-to-performance synthesis. It produced rule-based control of the amplitude, duration, formant frequencies and bandwidths of tones generated by the MUSSE synthesizer.

This possibility, to automatically convert a music file into a sounding performance caught the attention of Lars Frydén (1927 – 2001), well known violin soloist, string quartet head and music performance coach. He suggested that it would be interesting to have MUSSE sound as an instrument and have it play music files according to performance rules. Thus, our goal was to find out if it would be possible to formulate a system of ordered, context-dependent performance rules that would make the synthesizer play input music files as a reasonably gifted musician or at least as a gifted beginner.

The experimental setup is shown in Figure 10. The strategy was to process the input music file by a rule system, which contained a collection of context-dependent performance rules that controlled the MUSSE synthesiser. Frydén's expertise, resulting from his analytic mind combined with an extensive experience as a coach for chamber music ensembles, made it easy for him to indicate what changes were needed to improve a given performance of a given melody. His suggestions were given in a form that could be translated into performance rules. The rules were developed one by one, and for each rule, we carefully evaluated its effect on synthesised performances of melody examples. This procedure allowed us to spot deficiencies in terms of odd-sounding performance details and to modify the rule accordingly. Also, it was easy to identify performance rules that caused a problem and reformulate or replace it.

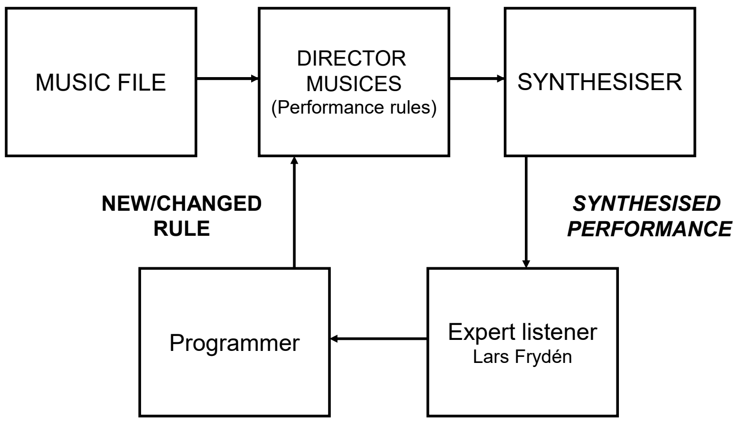

Fig. 10. Block diagram illustrating how the analysis by synthesis strategy was applied in the music performance research

The first description of the performance grammar was presented at a conference in Tallinn (Frydén, Sundberg, & Askenfelt, 1982). After that, several students and researchers participated in developing the system. Anders Friberg developed and organised it in his doctoral dissertation (Friberg, 1995). He named the system Director Musices, a freeware which can be downloaded at `https://www.speech.kth.se/music/performance/download/`.

Figure 11 shows the display of the Director Musices software, in this case provided with 24 performance rules. Each rule is represented by a rectangular window with a slider that can be dragged left or right; its position controls the magnitude (K) of the rule's effect, the numerical value of which is also displayed in the windows left of the sliders. If positioned in the middle, K is zero, so the rule is not applied, and pushing it right increases the effect. If positioned left of the middle K becomes negative, such that the rule is inverted.

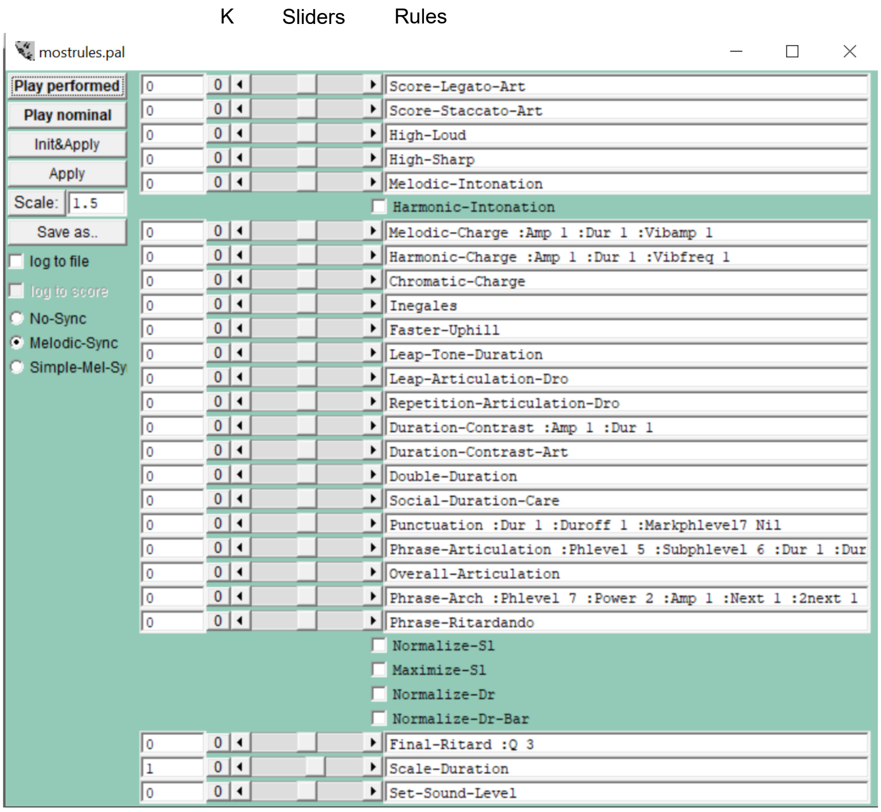

Fig. 11. Director Musices display with including 24 performance rules. The effect of each rule can be varied by changing the K parameter, which is controlled by the corresponding slider. Its value appear to the left of the slider.

Sound examples 2–6 demonstrate the effects of some rules. The rule *Duration contrast* states that short notes are played shorter and long notes are played longer than specified by the score, as illustrated in Figure 12. Thus, the rule increases the duration contrast between different note values. The first version of this Haydn theme (Sound example 2) is dead-pan, in the second version, the rule effect is exaggerated and in the third version it is more moderate (K=2.2). The fourth version with K=-2.2, thus lengthening short notes and shortening long notes, generates a quite odd-sounding performance.

Sound example 3 presents the *Pitch contrast* rule, which applies the principle "The higher the pitch, the sharper the intonation", thus approaching the

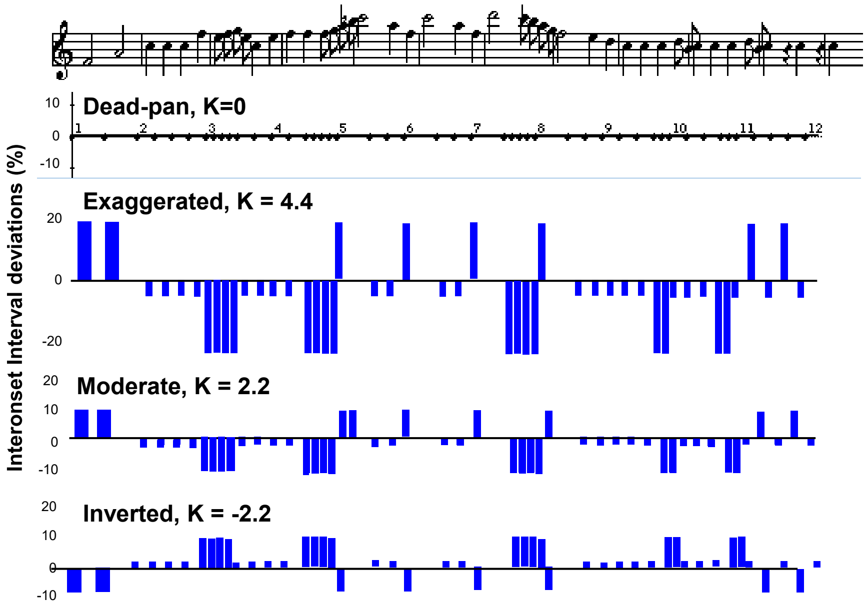

Fig. 12. Effect of four K values for the rule *Duration contrast*, which shortens short notes and vice versa, here applied to a theme from one of Haydn's string quartets

tuning of pianos. As in the preceding example the first version is dead-pan, the second exaggerated, the third moderate. In the fourth version the K value of the moderate version is multiplied by -1 and hence negative, which again produces a quite odd-sounding performance, in sharp contrast to the preceding moderate version.

Sound example 4 illustrates the *Melodic charge* rule, reflecting Lars Frydén's idea that different scale tones differ with respect to their remarkableness in the melodical context. The basic idea is that the further away along the circle of fifths that a tone lies from the root of the prevailing chord, the more remarkable it is in its melodical context and the higher its melodic charge. This is illustrated in the left part of Figure 13. Note that melodic charge is not symmetrically distributed around the circle of fifths; tones on the anti-clockwise (subdominant) side have negative charges.

The *Melodic charge* affects three tone properties; a tone's duration, amplitude and the extent of the vibrato are all proportional to its melodic charge, as illustrated in the right part of the figure and in Sound Example 4, the fugue theme in the first *Kyrie* of JS Bach's *B minor Mass.* The example is arranged as in the preceding examples, deadpan, exaggerated, moderate and inverted. Melodic charge is high and the effect is quite strong for the tone G in bars 3 and 5.

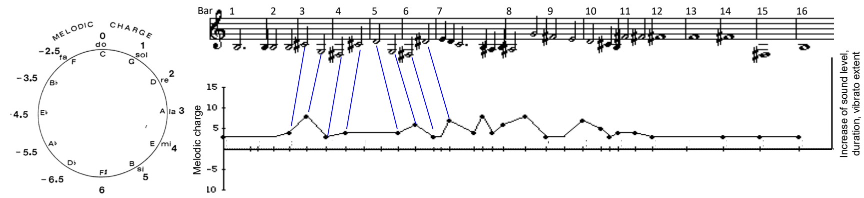

Fig. 13. Left: Melodic charge of the scale tones, arranged along the circle of fifths. Right: Effect of the rule in terms of increasing the loudness, duration and vibrato extent of tones according to their melodic charge.

Harmonic charge is a cousin of the *Melodic charge*, reflecting how remarkable a chord is in its harmonic context. For example, the dominant of the dominant is much more remarkable and consequently has a higher harmonic charge than the dominant. The rule is illustrated in Figure 14 and Sound example 5. The rules says that the amplitudes of tones are dependent on the magnitude of the change of harmonic charge in a chord progression; an increase is accompanied by a crescendo and a decrease by a diminuendo. The theme in the example, from the second theme in the first movement of Franz Schubert's *Unfinished symphony*, contains a remarkable chord shift; the theme is in D major, but in bar 6 a D♯ appears accompanied by a highly remarkable change of harmonic charge. As in the preceding examples, the first version is dead-pan, the second with a moderate quantity, while the rule is inverted in the third version.

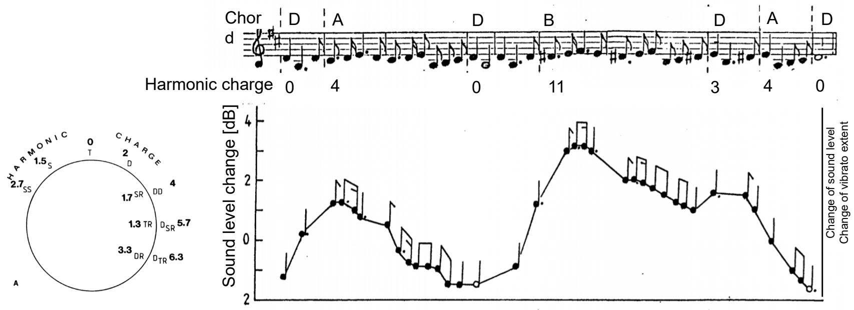

Fig. 14. Left: Harmonic charge of harmonies; T=Tonic, D=Dominant, S=Subdominant, R=parallel, so D_{TR} refers to the Dominant of the Parallel to the Tonic. Right: Derivation of crescendos and diminuendos from changes in harmonic charge in the second theme in the first movement of F. Schubert's "Unfinished" symphony. Minus sign after chord symbol denotes minor chord.

Sound example 6 demonstrates the effect of the rule *Phrase Arch*, probably one of the most important rules in performance of western traditional music. It accelerates the tempo in the beginning and decelerates it toward the end of phrases and sub-phrases, see Figure 15. The effect is greater for phrases than for sub-phrases. In other words, the rule produces a marking of constituents, enhancing the hierarchical structure of the piece in terms of small tempo changes.

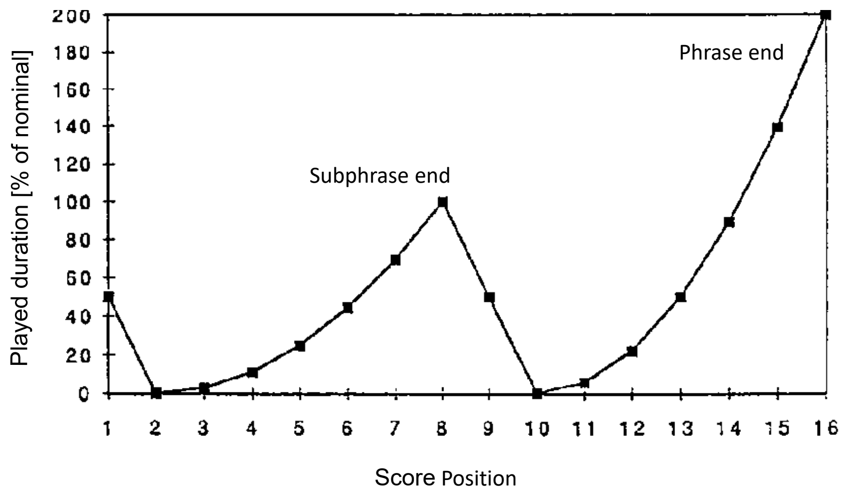

Fig. 15. Effect on tempo of the rule *Phrase Arch*

The performance rules contained in the Director Musices software may seem somewhat ad hoc. However, it can be argued that they constitute a description of the huge capital of musical experience and wisdom of a professional musician and music coach. Lars Frydén's expertise was to guide musicians to improved performances and he used the same skill when working with Director Musices. A typical situation: he found that a particular tone should not be played like that in this musical context, because it is much more remarkable than its neighbouring notes. He then formulated a hypothetical specification of the context and of how it should be played. His hypothesis could then be tested and fine-tuned in the analysis-by-synthesis setting.

4 Singing

Synthesis of singing is a project that was actually started before the music performance project, and it used the same equipment as that project. However, the timbre of the synthesis sounds was not critical in the latter project, while in synthesising singing a natural-sounding voice timbre is mandatory; a fair evaluation of a sung performance is almost impossible, if the synthesis fails to sound as produced by a living vocal artist. For this reason, the technicalities of the MUSSE synthesizer were more relevant in this application of the analysis-by-synthesis strategy and will be described first. A detailed technical description can be found in Larsson (1977).

In the department, in which I was working, speech synthesis was a main activity. My department colleague and close friend, the late Jan Gauffin, thought that synthesising singing would a valuable complement. One of his basic ideas was that the formant filters should be controlled by analogue signals from light-emitting diodes, thus avoiding stepwise changes in quickly changing synthesis parameters. An opportunity to realise this idea appeared in terms of the exceptionally gifted student, Björn Larsson. A keyboard offered the possibility to manual control of timing and pitch and all main parameters could be controlled with knobs. Tuning was pre-set to equally temperament. Vibrato rate, extent, and regularity could be set separately by knobs, and pitch transitions could be smoothed in various ways. There were five formant circuits, all individually variable with regard to frequency and bandwidth. The voice source was a pulse train producing a spectrum with an envelope slope of -6 dB per octave.

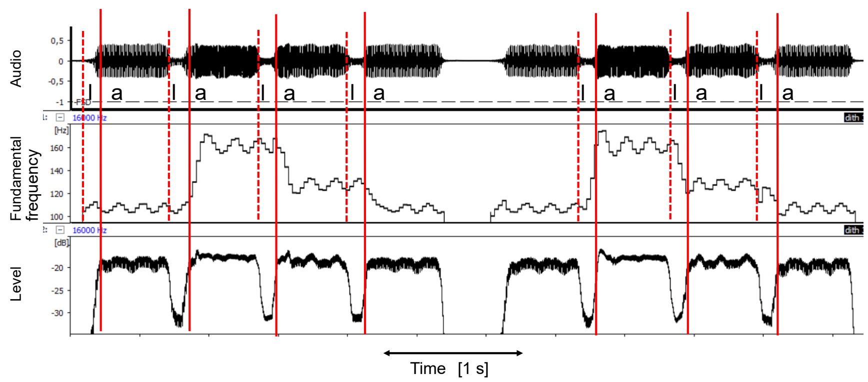

Fig. 16. Fundamental frequency patterns obtained from synthesis of a quick, legato sequence of short notes, performed on /la la la la/ without (left) and with (right) application of *Coloratura* rule. Dotted and solid lines mark onsets of the pitch change and the vowel, respectively

A second major step in the development MUSSE was an interface offering the possibility of synthesising also voiceless consonants. A third step in the development was that Rolf Carlson and Björn Granström modified their RULSYS software (Carlson & Granström, 1975, 1976; Carlson, Granström, & Hunnicutt, 1982). As mentioned, it operated with a system of ordered, context dependent rules. The modification generated a musical cousin of the RULSYS text-to-speech system, a quasi-score-to-tone system. In its new, musical garment, the RULSYS rule-system could be loaded with ordered, context dependent rules for sung performance. One of the first presentations of MUSSE synthesis happened at an IRCAM symposium about fifty years ago.

Sound example 7 contains three versions of a *Vocalise* by Heinrich Panofka from his The Art Of Singing, op.81. It illustrates the possibility to analyse detailed aspects of expressiveness in singing. The first version is dead-pan. In the second version one single performance rule is introduced. A fundamental frequency undershoot is added to the onset of each note, which generates a marcato effect, perhaps more appropriate for a military march, and certainly alien to this composition. In the third version a rule marking harmonic charge has been added. It thus gradually increases the amplitude toward shifts to chords more distant from the tonic than the preceding chord. In the example, a manual piano accompaniment played by a pianist was added.

Actually, I made a little joke with this. Under a falsified name, I sent the last version in this example to a choir leader and asked if it would be worthwhile for me to join the choir? I was greatly encouraged to note that the choir director needed no less than two weeks to decide, but then answering "In my choir you must be capable of singing also consonants".

The work with synthesis of singing attracted attention to several characteristics of sung performance that were unexpectedly relevant. Sound example 8 and Figure 16 provide an example. The question was: Where is the correct place for a voiced consonant? In the first version, the pitch change happens during the vowel onset, while in the second it occurs during the consonant. Each version occurs three times. The synthesis corroborates the general opinion among choir leaders that the second alternative is the correct one.

Performance of *legato* sequences of short notes, or *coloratura*, is illustrated in Sound example 9 and Figure 17. Again, the first version is dead-pan, producing an unclear articulation of the descending part of the pitch pattern. Also, the top note in each ascending-descending sequence sounds somewhat flat. In the second version, part of the *Marcato* rule mentioned earlier is applied; each tone is approached from one semitone below the target. In the middle of the tones' durations the fundamental frequency is set to two semitones above the nominal target. In addition, the mid-tone target of the top note in the ascending-descending melody is sharpened by three semitones. After the smoothing of the quasi square-wave fundamental frequency contour with a low pass filter, the pitch of each note is more clearly articulated, and the top note does not sound flat.

Sound examples 10 and 11 present two examples demonstrating the quality that was eventually reached in this project. In Example 10 MUSSE is dressed up as a tenor chorister singing the theme of the first *Kyrie* in J.S. Bach's *B minor mass.* The double bass and keyboard accompaniment was played by live musicians and added.

The G4 in bar 8 (see Figure 13) demonstrates an important deficiency of this synthesis. The formant frequencies of the vowel /y/ in the word "Kyrie" were kept constant, irrespective of pitch. This is atypical in singing; when the pitch frequency is higher than the normal value of the first formant, singers

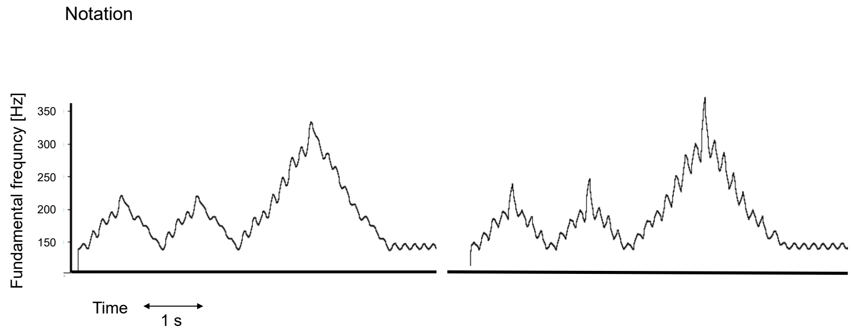

Fig. 17. Fundamental frequency patterns obtained from synthesis of a quick, legato sequence of short notes, performed without (left) and with (right) application of *Coloratura rule.*

typically raise the first formant to a frequency somewhat higher than the pitch frequency (Sundberg, 1975). The pitch frequency of G4 is 392 Hz, while the first formant of the /y/ was set to 330 Hz. As a result the synthesized singer voice sounds somewhat unnatural or "strangled" on that pitch. It would obviously have been simple to introduce the rule that would eliminate this.

Again, I made a little joke with this example. I sent it, again under faked name, to a US university, asking if my singing was good enough for me to be accepted for a doctor of music education at the university. I was greatly encouraged that the answer was affirmative.

Sound example 11, finally, presents a more entertaining case, demonstrating the possibility to extend the pitch range of a synthesised voice far beyond the limits of a real singer. Composer Gerald Bennett kindly agreed to composed a piece, *Limericks*, for the final *Voice Synthesis* session of the *Stockholm Music Acoustic Conference (SMAC) 1993* (Bennett, 1993). The synthesis of the final verse was made by Gunilla Berndtsson.

5 Outlook

Above, three research projects have been described in which the analysis-by-synthesis method has been applied to research in the area of music science, (1) description of a compositional style, (2) the gap between what is written in the score and the sounding realization of a score, and (3) sung performance. A common denominator of these investigations is that they provide an answer to the question. "Is your description of the research object exhaustive?" Generally, odd-sounding details implies that the answer is negative. Only when the description has been developed, perceptual testing of results with expert panels is appropriate (e.g., Thompson, Sundberg, Friberg, & Frydén, 1989).

In the generative grammar of Tegnér's nursery tunes, the hierarchical structure turned out to be very productive. When transformed into a prominence contour, it served as an entrance ticket to the rules for tone duration, chord progression and pitch assignment. Hierarchical structure was a key concept also in the musical performance grammar, suggesting that marking phrase and subphrase boundaries e.g., with tempo changes, is a quite dominant principle in music performance. This corroborates the observation that hierarchical structure is quite important in traditional music (see e.g., Todd, 1985).

Selecting the compositional style of nursery tune certainly simplified the task to formulate a generative grammar for describing it; the structure, the metrics, the harmonic progressions and the melodic lines were all uncomplicated, presumably tailored for young minds. Still, analysis-by-synthesis has been applied also to more complex styles. It was used in an attempt to define the similarity between of a great number of variants of a Swedish folk tune, the *Fiskeskärsvisa*, and the success was demonstrated in generating new variants by randomised choice between alternatives allowed by the rule in the grammar (Sundberg & Lindblom, 1976). Likewise, analysis-by-synthesis has been applied to performance of electroacoustic music (Friberg, Frydén, Bodin, & Sundberg, 1988).

Analysis-by-synthesis often facilitates identifying reasons for odd-sounding events. It is often easy to go back to the rule system and spot what is causing the problem. As mentioned, our first attempts to synthesise nursery tunes clearly demonstrated the importance of a pitch proximity principle in terms of frequent occurrence of wide intervals between adjacent tones. This was a quite salient problem that was easily solved by the *Pitch adjacency* rule; when attributing a pitch to a tone that is located between two tones with already determined pitches, only pitches located between these pitches are allowed.

In some cases, odd-sounding events provoked identification and quantification of as yet less well-defined characteristics. For example, it was obvious that different scale tones needed to be played differently, depending on how *remarkable* they were in their musical context. We tried many ideas, among which introducing the Melodic charge concept produced the best results. As mentioned, it reflects the scale tone's distance along the circle of fifths from the root of the prevailing chord. This may seem surprising, as it does not differentiate chord note and non-chord notes, e.g., suspensions. On the other hand, most non-chord notes have a higher melodic charge than their neighbour chord note. Melodic charge was a key concept also for Harmonic charge values to chords, defined as a weighted sum of the chord tones' melodic charges. It generates dynamic variations reflecting chord progressions. Both Melodic charge and Harmonic charge may appear as ad hoc inventions. However, they are both strongly correlated with results from independent experiments; both have a correlation of 0.78 with ratings of degree of "good continuation" that Krumhansl and associates obtained in a series of experiments, where a probe tone or a probe chord was presented after a melodical and a harmonic context, respectively (see e.g., Krumhansl, 2001).

An unexpected experience emerged during work with tuning rules in the performance grammar was that surprisingly small effects could be perceptually quite important. For instance, the difference between two performance versions could be clearly heard in paired-stimulus tests, even though it was impossible to tell what the difference was. Detectability of minute disturbance of duration is amazing; for tones shorter than 240 ms, presented in an isochoronous six-tone sequence, the just noticeable change of the duration of one of the tones is no more than 6 ms (Friberg, 1995). It is thought-provoking that skilled ensembles can keep synchrony also in sequences of such short tones, in spite of the fact that a conductor generally marks only the timing of the beats.

The application of analysis-by-synthesis to music also seems to shed some light on music communication in general. Hierarchical structure was found to play a key role in both composition and performance of the tonal Western music that we analysed. This suggests that both composer and performer strive to help the listener to organize the tonal material as a hierarchical structure of chunks.

Many of the rules in the synthesis of musical performance seem to serve either of two purposes, *grouping of sound events* and *differentiation of tone categories*. For example, the changes of tempo and dynamics that mark phrase boundaries would facilitate grouping, and the emphasising of tones with high melodic charge is a case of differentiating scale tones.

As mentioned, a recurrent experience has been that even vanishingly small deviations from deadpan performances may be quite important. This suggests that the performance grammar can be used as a tool for training analytic listening. In addition, it suggests that in top quality performances random deviations from deadpan are likely to be noticeable and thus unlikely to occur.

Finally, it might be mentioned that negative experiences when presenting results obtained from analysis-by-synthesis work do not necessarily need to be discouraging. For instance, when Björn Lindblom and I first presented the compositional grammar at a musicology conference, the experts' reaction was quite emotional. One delegate's comment was "These are the worst nursery tune compositions I have ever heard!" We had no problem with agreeing with this evaluation, but felt no reason to be discouraged, knowing that analysis-by-synthesis offers excellent possibilities to identify and solve remaining problems. This is true, no matter if the research object is compositional style, music performance or singing.

6 Conclusions

Combining these examples of applying analysis-by-synthesis to compositional style, to music performance and to the singing voice supports the following conclusions about basic aspects of music communication.

Deviations from nominal score are crucial in musical performance, and even vanishingly small deviations can be important.

Grouping, i.e., showing what events belong together and where structural boundaries occur, seems a basic principle in both composition and performance.

Differentiation of tone categories, e.g., by exaggerating the difference between them, or by emphasising remarkable events and de-emphasising expected events, seems a basic principle in music performance.

Analysis-by-synthesis can help the identification and quantitative description of perceptually important properties, which earlier were only intuitively known and referred to in unclear terms.

Acknowledgements The material in this article was first presented as a key note at the 2020 Joint Conference on AI Music Creativity. The permission of the journal *Advances in Cognitive Psychology* to use figures and sound examples from Sundberg (2006) is gratefully acknowledged. Susanna Sundberg kindly transcribed the recording of the presentation.

Sound Examples

Sound example 1 Bengt Hallberg playing his arrangements of six rule generated tunes.

Sound example 2 Demonstration of the rule *Duration contrast*, which shortens short note values and lengthens long note values, see Figure 12. Four versions: (1) Deadpan K=0, (1) Exaggerated K=4.4, (1) Moderate, K=2.2, (1) Inverted, K=-2.2. Link to sound example.

Sound example 3 Demonstration of the rule *Pitch contrast*, which sharpens high pitched notes and flattens low pitched notes. Four versions: (1) Deadpan K=0, (2) Exaggerated K=4.4, (3) Moderate, K=2.2, (4) Inverted, K=-2.2. Link to sound example.

Sound example 4 Demonstration of the rule *Melodic charge*, which emphasizes tones with vibrato and loudness depending on their melodic charge. The example is the theme of the first *Kyrie* in J.S. Bach's *B Minor Mass*, first performed deadpan, then with the rule applied in moderate quantity. See Figure 13. Link to sound example.

Sound example 5 Demonstration of the rule *Harmonic charge*, which slows the tempo when an increase of the harmonic charge is approaching and vice versa. The example is the second theme in the first movement in F. Schubert's *"Unfinished" symphony*, first performed deadpan, then with the rule applied in moderate quantity, and last with the same rule applied inverted. See Figure 14. Link to sound example.

Sound example 6 Demonstration of the rule *Phrase Arch*, which accelerates tempo in the beginning of a phrase and slows it in the end of a phrase, see Figure 15 and 16. Four versions: (1) Deadpan K=0, (2) Exaggerated K=2.5, (3) Moderate, K=2.2, (4) Inverted, K=-2.2. Link to sound example.

Sound example 7 Dead-pan and processed versions of MUSSE performance of a Vocalise from Heinrich Panofka's The Art of Singing, op.81. (From Sundberg, 1975). Sound example in (Sundberg, 2006).

Sound example 8 Timing of pitch change. In the first version, the pitch change is located during the first part of the vowel, in the second version it is located in the consonant. Each version is played three times (From Thompson et al., 1989). Sound example in (Sundberg, 2006).

Sound example 9 Coloratura. Two versions of a sequence of short notes, the first without and the second with application of the coloratura rule, which makes the fundamental frequency turn around each target frequency, see Figure 4 (From Thompson et al., 1989).Sound example in (Sundberg, 2006).

Sound example 10 Synthesis of the first theme of the first Kyrie from J S Bach's B Minor Mass. The double bass and keyboard accompaniment, played by two musicians, was added to the synthesis.Sound example in (Sundberg, 2006).

Sound example 11 Last verse with refrain from Gerald Bennett's Limericks, composed for the International Ensemble of Synthesized Singers (IESS). Lyrics:

(Solo) There was a young boy in the choiyah
whose voice rose highyah and highyah,
till one balmy night
it shot clear out of sight.
They found it next day in the spiyah.

(Chorus) Will you come up to Limerick,
will you come up to Limerick?

The voice was synthesized on the MUSSE machine and the piano on a computer controlled synthesizer (From Berndtsson & Sundberg, 1994).

References

Bennett, G. (1993). Limericks in CD record. In *SMAC93, Proceedings of the Stockholm Music Acoustics Conference.*

Berndtsson, G., & Sundberg, J. (1994). The musse dig singing synthesis. In *Smac 93 (proceedings of the stockholm music acoustics conference 1993), stockholm, roy. sw. academy of music, publ* (pp. 279–281).

Carlson, R., & Granström, B. (1975). A text-to-speech system based on a phonetically oriented programming. *Language, Speech Transmission Laboratory, KTH Stockholm, Quarterly Progress and Status Report*, 17–26.

Carlson, R., & Granström, B. (1976). A text-to-speech system based entirely on rules. In *Icassp'76. ieee international conference on acoustics, speech, and signal processing* (Vol. 1, pp. 686–688).

Carlson, R., Granström, B., & Hunnicutt, S. (1982). A multi-language text-to-speech module. In *Icassp'82. ieee international conference on acoustics, speech, and signal processing* (Vol. 7, pp. 1604–1607).

Friberg, A. (1995). A quantitative rule system for musical performance. doctoral dissertation, speech music and hearing. *http://www. speech. kth. se/music/publications/thesisaf/sammfa2nd. htm..*

Friberg, A., Frydén, L., Bodin, L., & Sundberg, J. (1988). Performance rules for computer music. In *Atti, VBII Colloquio di Informatica Musicale.*

Frydén, L., Sundberg, J., & Askenfelt, A. (1982). A rule system for converting melodies from musical notation into sound. In *Proceedings of Symposium on Common Aspects of Linguistic and Musical Data.*

Gabrielsson, A. (2003). Music performance research at the millennium. *Psychology of music*, *31*(3), 221–272.

Gauffin, J., & Sundberg, J. (1974). Masking effects of ones own voice. *Speech Transmission Laboratory Quarterly Progress and Status Report (KTH, Stocholm)*, *1*, 35–41.

Krumhansl, C. L. (2001). *Cognitive foundations of musical pitch* (Vol. 17). Oxford University Press.

Larsson, B. (1977). Music and singing synthesis equipment (musse). *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)*, *1*(1977), 38–40.

Seashore, C. E. (1967). *Psychology of music.* Dover publ.

Sundberg, J. (1975). Formant technique in a professional female singer. *Acta Acustica united with Acustica*, *32*(2), 89–96.

Sundberg, J. (2006). The kth synthesis of singing. *Advances in cognitive Psychology*, *2*(2), 131.

Sundberg, J., Askenfelt, A., & Frydén, L. (1983). Musical performance: A synthesis-by-rule approach. *Computer Music Journal*, *7*(1), 37–43.

Sundberg, J., & Lindblom, B. (1976). Generative theories in language and music descriptions. *Cognition*, *4*(1), 99–122.

Tegnér, A. (1892). *Sjung med oss, mamma!* Nordiska musikförl.

Thompson, W., Sundberg, J., Friberg, A., & Frydén, L. (1989). Rules for expression in the performance of melodies. *Psychology of Music*, *17*(1), 63–82.

Todd, N. (1985). A model of expressive timing in tonal music. *Music perception*, *3*(1), 33–57.