Musician’s tone glue

Sundberg, J.

journal: STL-QPSR
volume: 33
number: 1
year: 1992
pages: 079-098

http://www.speech.kth.se/qpsr
MUSICIAN'S TONE GLUE

Johan Sundberg

Abstract

In music performance, musicians rely on a variety of methods to mark melodical conherence of notes, i.e., that the notes belong together and constitute a melodical Gestalt. Minute departures from the nominal duration of the tone, as given by the score, seem to be particularly efficient but also loudness and timbre are used. Our generative rules for music performance (A. Friberg, Computer Music J. 15, pp. 56-71, 1991) offer some examples: the final tone of a phrase is lengthened, and subphrases are terminated by a micropauses. Likewise, a crescendo seems to have the effect of joining the tones involved. A change of timbre seems another efficient method to signal a boundary between tone groups. The improvised nonsense texts that we use when we sing tunes also present examples of musical grouping. The magnitude of the effects needed to elicit a musically useful effect is discussed.

THE PROBLEM

A deadpan performance of a piece of music seems far from acceptable from a musical point of view. It is as if one could hear, with a painful evidence, how little the player cares about the piece he is playing. He ignores phrases and articulation, and he treats important and unimportant tones alike. When one listens to such performances, the piece appears to consist of an endless, nagging sequence of tones that are completely independent of one another.

What is the problem, then? What we seem to miss is the marking of harmonic progressions, of phrases, of exciting melodic leaps. In a deadpan performance, the player's comments on what is being played is all but missing, and this is a deficiency that irritates in the same way as it is irritating to listen to someone apparently void of compassion read a poem. The result is that the piece of music disintegrates into small fragments, and sounds boring.

This is probably a notorious experience for music teachers. In our time, the problem has been accentuated by the advent of playing machines, such as computers, sequences, synthesizers, etc. It is now easy to produce music that is played exactly as written in the notation. A typical experience seems to be that such verbatim performances are boring, no matter how interesting the composition is.

METHOD

This observation caused the violinist and music teacher Lars Frydén and myself to start a research project in the end of the seventies. The construction of a singing machine, MUSSE (Music and Singing Synthesis Equipment), had just been successfully completed (Larsson, 1977). It could be played from a keyboard. The vowel timbre was almost flawless, but whatever happened musically in the composition, the machine invariably sounded completely blank. The lack of musical engagement turned out to be a musical disaster.

In his work as a music pedagogue, Lars Frydén was curious to find out if his instructions to students were to be understood literally or artistically. He also asked himself why it was necessary to do all these small tricks when playing? Was it a
question of rules that should always be applied, as some teachers tend to pretend? Could a computer be programmed such that it could replace a musician?

The possibility to attack this problem was very prosperous at that time. An interface was built so that the singing synthesizer MUSSE could be played from a computer. Rolf Carlson and Björn Granström had constructed a computer program that converted written text in an input file to sounding speech (Carlson & Granström, 1975). They modified their text-to-speech conversion program to a note-to-tone conversion program, so that MUSSE could convert an input note file to singing. In their text-to-speech program, there were good provisions for implementation of context-dependent pronunciation rules and by listening to the result in terms of synthesized speech, one could hear if the rules were appropriate or not. Similarly, we could implement context-dependent performance rules in the music application. Thus, we could have an automatic, rule-controlled conversion of an input note file to a sounding performance, and we could try different rule formulations and listen to the effects it had on the performance.

This research paradigm is well-known within the natural sciences. It is called analysis-by-synthesis. One analyzes, e.g., sound, by synthesizing it. In this way, it becomes possible to experience the significance of different acoustic characteristics from an auditory point of view.

This equipment proved to be an almost perfect tool for analyzing music performance. As the sound quality could easily be adjusted so that it sounded almost exactly as a living singer or as a wind instrument, one expected realistic performances from the synthesis. It was possible to implement various context-dependent performance rules that introduced lengthenings, accents, crescendos, etc. in the performance, and then listen to the musical effect. During most of the time, Lars Fryden and I have worked together with Anders Friberg, pianist and computer programmer. In this article, I will present some rules that have emerged from this research (Sundberg, Friberg, & Fryden, 1991a). Also, some other research results from investigations of music performance will be reviewed.

PERFORMANCE RULES

The complete set of our performance rules has been presented in detail elsewhere (Friberg, 1991). Table I presents an overview in which the rules have been grouped according to the purposes they appear to serve in the music communication process. We have identified three such purposes. One is to keep order within an ensemble, i.e., to decide when the tones shall start and end, and how intervals and chords should be tuned. These rules are called ensemble rules.

Another purpose seems to be to facilitate the listener to correct differentiation between tone categories, both the pitches into the scale tone categories and the durations into the note value categories. By means of these differentiation rules, the player helps the listener to realize if a tone is remarkable or trivial in its melodic context, if it is an eighth note or quarter note, etc. The correct categorization seems important for the listener's understanding of which tones belong together.

A third group of rules, the grouping rules, seems to more directly serve this last-mentioned purpose, to facilitate the listener's grouping of tones, thus relating closely to the main topic of this article. These rules help the listener to realize which tones belong together, and where the boundaries are in the musical structure.
Fig. 1. Illustration of a hierarchical structure in a nursery tune by the Swedish composer Alice Tegnér.
### SEGMENTATION RULES

**1. Duration categories**
- **DDC 1A** - Durational Contrast
- **DDC 1B** - Durational Contrast
- **DDC 2A** - Accents
- **DDC 2B** - Double Duration

**2. Pitch categories**
- **DPC 1A** - High Sharp
- **DPC 1B** - High Loud
- **DPC 2A** - Melodic Charge
- **DPC 2B** - Melodic Intonation

### GROUPING RULES

**1. Micro level**
- **GMI 1A** - Leap Articulation
- **GMI 1A’** - Same, alternative formulation
- **GMI 1B** - Leap Tone Duration
- **GMI 1C** - Faster Uphill
- **GMI 2** - Amplitude Smoothing
- **GMI 3** - Inegalles
- **GMI 4** - Articulation in Repetition
- **GMI 4’** - Same, alternative formulation

**2. Macro level**
- **GMA 1** - Phrase
- **GMA 2A** - Harmonic Charge
- **GMA 2B** - Chromatic Charge*
- **GMA 3** - Final Ritard

### ENSEMBLE PERFORMANCE RULES

- **ENS 1** - Mixed Intonation
- **ENS 2** - Melodic Synchronization
- **ENS 3** - Bar Synchronization

*Table I. Overview of the performance rules affecting duration (DR), sound level (L), fine tuning (F), vibrato amplitude (VA), and vibrato frequency (VF) or inserting micropauses at the end of notes (DRO). Some rules have two alternative (alt.) formulations.*

### Hierarchical structure

Almost all music is built according to a *hierarchical principle*. This principle, illustrated in Fig. 1, implies that sets of adjacent, small units or *constituents* form intermediate constituents, which together with other intermediate constituents form greater constituents, which together with other greater constituents form still greater constituents, which together with other such constituents form *still* greater constituents, etc.
The hierarchical structure seems to be somewhat like a finger print or rather, perhaps, as a brain print on most creations of the human intellect. Government is a typical example of a hierarchical structure, all the way from president or king and down to the cleaning personnel. Classical architecture often offers other striking examples of hierarchical structures, starting with the entire house, via the center part and the wings, via roof, walls, and souterrain, via windows, doors, and wall areas and down to the details in terms of bricks etc. Indeed, the human mind seems to have difficulties in handling structures that are not hierarchically organized.

Hierarchical structure can be found also in music. Figure 1 illustrates the very obvious hierarchical structure in a simple piece of music that, typically enough, is written for children. This structure is representative for 16-bar nursery tunes in 4/4 time composed by the Swedish composer Alice Tegnér (Lindblom & Sundberg, 1970; Sundberg & Lindblom, 1976). Mostly, the hierarchical structure is less symmetrical and perhaps a little less obvious in other types of music. However, the hierarchical principle is applicable to the structure of almost all types of music. Tree diagrams, like that shown in Fig. 1, offer a means to efficiently represent the structure (Lerdahl & Jackendoff, 1983).

The hierarchical structure is the key to many aspects of music performance. Out of respect for this principle, let us start from the bottom and work our way upward.

Grouping rules
The smallest constituent in the musical structure is the single note. But notes belong together in different ways. If they appear as a scale, this scale is often a constituent. According to one of our performance rules, a tone initiating an ascending interval is shortened. In the lab jargon, this rule is called Faster Uphill, because it increases the tempo in ascending scales. Most listeners agree that this rule has the effect of giving the scale a direction of movement, gluing together the tones in ascending scales so that they form a Gestalt. A grouping effect emerges.

If all tones in a piece are played as a never-ceasing, invariably applied legato, the performance tends to sound boring. Thus, the minute micropauses that players generally insert in the tone flow are important. Another performance rule, Leap Articulation, inserts micropauses in melodic leaps. Physically speaking, it makes silent the last portion of all tones that initiate leaps, and the duration of this silence is proportional to the width of the interval as measured in semitones. Most listeners agree that the insertion of such pauses improves the musical quality (Sundberg, Friberg, & Frydén, 1991b).

Why do music listeners prefer these micropauses? It seems reasonable to assume that, as a general principle, pauses mark boundaries in the structure. In any event this is true for speech. When, while speaking, we arrive at a comma or period, we generally insert micropauses, and we also tend to lengthen the last syllable before a period.

This leads us to a hypothesis that may appear as rather bold: composers of traditional Western music often tend to introduce leaps at structural boundaries while narrow intervals are more appropriate to mark continuity. There is some support from other research for this hypothesis.

Psychoacoustic experiments carried out by the Dutch psychologist Leon van Noorden (1975) is one. He presented series of equally short tones to listeners and varied both the durations of the tones and the interval size between adjacent tones.
He found that for wide intervals, the melodical line broke into two interleaved, autonomous tone sequences, one consisting of the sequence of high tones and another consisting of the sequence of low tones. When the intervals between adjacent tones were made smaller, the two lines fused into one line. The critical interval, at which the two melodic lines started to merge, depended on the duration of the notes. A melodic line could survive wide melodic intervals provided the tones were long, i.e., if the tempo was not too quick. In this sense, these results support the above assumption that melodical Gestalts are likely to occur in scale sequences, because the intervals between adjacent tones are small in this case, and, by contrast, structural boundaries are likely to occur at wide leaps.

Further support for this same hypothesis, that tones in scale sequences belong together and form constituents, and that melodic leaps occur at structural boundaries, can be found in the classical Gestalt laws. The laws Similarity and Proximity both seem applicable, contributing to the effect that small intervals tend to glue tones together. Such tones are similar in pitch, as they are close in frequency. They also possess proximity both in time, if they are adjacent, and in frequency, if the intervals between them are narrow, such as in a scale.

A third reason to believe in this hypothesis can be found in a very interesting contribution to music theory by Simon & Sumner (1968). They claimed that the letters in an alphabetical enumeration tend to stick together, forming constituents. Further, they claim that the musical scale can be regarded as an equivalent of the alphabet. This supports the idea that scale sequences tend to stick together, forming structural constituents and, consequently, leaps mark structural boundaries.

Singing
The above rules can be successfully applied also in singing (Sundberg, 1989). However, some specific performance rules can be observed in singing. When a singer practices a new song he/she sometimes finds it helpful to skip the consonants and sing the music on just one or a few vowels. For pedagogical purposes, singing teachers have composed vocal exercises in terms of songs without text that are supposed to be sung on vowels only. Such songs are called vocalises. In sung songs, the disturbances of the vocal-fold function caused by, e.g., stop consonants can be avoided, and a smooth tone production is facilitated throughout the singer's entire pitch range. In such cases, there are no consonants that could help the listener to identify the boundaries between adjacent tones. Under these conditions singers often use the pitch gesture illustrated in Fig. 2; in descending intervals the pitch change is produced with an overshoot, so that the fundamental frequency descends too far and then rises again up to the target. The effect of this is a marcato. As has been demonstrated in a synthesis example published previously (Sundberg, 1989), this effect can be enhanced by small crescendo-diminuendo gestures on each note.

This method of marking the boundaries between tones seems logical. Great signal changes typically occur at tone boundaries, at least when the tone corresponds to a complete syllable. Then, there is often a consonant at the beginning of each tone. Consonants imply great changes in the acoustic characteristics of the tone produced, although apparently not great enough to necessarily generate the impression of a structural boundary. In vocalise singing, such changes are lacking, so the singer exaggerates the only change left, i.e., the fundamental frequency change. Does it happen less, when the consonants are sung.
Fig. 2. Fundamental frequency versus time for a singer performing a Vocalise, i.e., a song sung on vowels, without any consonants. Note the overshoots between tones in descending melodic motion. (After Sundberg, 1978.)

The rules discussed so far were discovered by means of the analysis-by-synthesis method, i.e., by listening to the computer controlled synthesizer's attempts to play music examples in a musically acceptable way. An advantage with the analysis-by-synthesis method is that one can tease out the function of single rules. For example, there are many but quite different reasons to lengthen a tone; a lengthening may mean phrase ending, emphasis and perhaps other things too. If an analysis-by-synthesis method is used, different cases of lengthening can be tried in terms of context-dependent, exploratory performance rules and the effect on the performance can be heard.

A more commonly used method in music-performance research has been analysis-by-measurements on real performances. Thus, one measures, e.g., the tone durations in a piece as played by different performers. One difficulty with this method is that it is difficult to identify the reason for, e.g., a lengthening of a particular tone, as suggested above. Furthermore, different performers may lengthen different tones, because their interpretations differ. Therefore, durations averaged over several performers may be quite difficult to interpret. Nevertheless, measurements are often, of course, quite revealing.

Measurements can be carried out not only on the sound produced but also on the input parameters, i.e., on the musician’s gestures. The latter is the case for the three examples shown in Fig. 3. They show the changes of lung pressure in a singer singing descending scales. The figure is quoted from a recent investigation (Sundberg, Gramming, & Elliot, 1991). When we speak or sing, we use this pressure to regulate vocal loudness; the higher the pressure, the louder the tone. In this case, lung pressure was captured as the esophageal pressure which faithfully reflects lung-pressure changes as long as the lung volume is constant. Thus, for adjacent tones, lung-pressure changes can be studied in terms of the changes of the esophageal pressure. The
esophageal pressure can easily be measured if the subject swallows a thin catheter with a minute pressure transducer.

Fig. 3. Relative lung pressure in a singer singing descending scales in 3/4 time (the two upper curves) and 4/4 time (bottom curve) on the syllable [la], repeating each scale tone three and four times, i.e., throughout each bar. In the case shown at the top, the singer was asked to avoid the marking of bars.

The figure shows the pressure events during performance of two types of scales, one in 3/4 time, the other in 4/4 time. In each case, each scale tone was repeated throughout the bar. Thus, in the 4/4 time scale, each scale tone was repeated four times in each bar, and in 3/4 time each scale tone was repeated three times. This figure reveals how this singer marked the bar unit which was obviously a prominent constituent in this very simple musical structure. The singer produced a short pressure accent on the first beat in each bar in both the 4/4 time and the 3/4 time scales. When he was asked to avoid marking the bars in the performance, he stopped producing such pressure accents, the pressure becoming very even, as can be observed on the top graph in the same figure.
The same experiment was carried out also with a female singer. She did not use
the same strategy. Her method was instead to produce a very gentle pressure rise
throughout each bar, culminating on the first beat in the next bar. Thus, she made a
slight crescendo during each bar, starting on the second beat. In addition, she used
articulatory means to mark the bar unit.

These examples indicate that the singer can use different means to mark the bar
unit in the musical structure, whenever he/she finds this desirable. One way is to
add a pressure accent to the first beat in each bar. Another method is to produce a
slight crescendo culminating on the first beat in the bars. In both these cases, how-
ever, the entering into a new structural unit was marked by a difference in loudness.

**Phrasing**

Let us now return to our interpretation rules developed by analysis-by-synthesis.
The structural constituent closest to the bar is the phrase. In the hierarchical structure
there seem to be at least two levels below the opening and closing phrases. They can
be called *subphrase* and *phrase*. An example of this can be seen in Fig. 1, showing a 16-
bar melody with subphrases and phrases marked. The structure is perfectly sym-
metrical; each phrase consists of two subphrases. This seems typical for music de-
dsigned for young brains: as mentioned, the structure pertains to a nursery tune.

According to the *phrasing* rule in our performance rule system, subphrases are
terminated by a micropause while the last note of the phrase is marked by a 40 ms
lengthening. This means that the subsequent tone is started a little later than would
have been the case otherwise.

Why are the phrases marked in this particular way? The answer probably is that
this is what we are accustomed to from speech. When we finish a sentence, we usu-
ally lengthen the last syllable, as mentioned. This is a well-known phenomenon in
phonetics where it is referred to as *final lengthening*. It seems to occur in a majority of
the languages in the world. Furthermore, we often insert a small pause at important
structural boundaries in speech. Indeed, this looks like a smart idea: music listeners
understand speech, so therefore let us use the same method of marking termination
and boundaries in music as in speech!

In traditional Western tonal music, harmonic progressions are frequently used to
mark the structure. For instance, the progression dominant-tonic (chords V-I) is often
used to signal the end of a structural constituent such as a phrase. These progress-
sions seem relevant to music performance. According to one of the performance
rules, *marking harmonic charge*, such progressions are marked with crescendos and
diminuendos. In short, the rule states that crescendos are introduced that culminate
on dominants, and diminuendos are used to announce the subsequent advent of the
tonic chord. Thus, crescendos seem to warn the listener that a remarkable chord is
about to arrive, and diminuendos signal the advent of unremarkable chords. In our
performance rule system, the harmonic charge of a chord is a quantitative estimate of
its remarkableness. This rule seems to serve the purpose of marking the structure;
composers use chord progressions to mark the structure and the chord rule marks
these chord progressions.

Crescendos and diminuendos also seem to serve another purpose in the marking
of structure, namely to glue together the tones involved in the crescendo or in the
diminuendo. A crescendo implies a gradual increase of loudness that is evenly dis-
tributed over a string of adjacent tones. In other words, these tones together are
united by jointly forming a change of the overall loudness. In a way, one may regard this function of the crescendo/diminuendo as iconic: it is as if the notes are telling the listener: So we all do the same thing because we belong to the same constituent.

**Timbral legato**

Singers spend much time trying to develop a beautiful legato, i.e., a seamless representation of a melodic sequence of tones. To achieve this goal, it seems important that the singer can combine wide pitch leaps with an unchanged vocal timbre; the voice quality should be the same even for tones that are far remote from each other with respect to pitch.

Nothing of this sort is required in speech where voice timbre typically varies with pitch in speech. For example, the larynx normally moves up and down with the voice pitch when we speak. The larynx position affects the voice timbre. This is because an elevation of the larynx shortens the vocal tract and therefore raises the formant frequencies, thus producing a "brighter" timbre. Conversely, a lowered larynx tends to make the vowel sound "darker" (Sundberg, 1987).

If the pitch-dependent larynx height were used also in singing, this would mean that tones differing considerably in pitch also would differ clearly in voice timbre. This would tend to break the legato line at wide pitch jumps. Therefore, it seems important for singers to learn how to timbrally bridge wide pitch gaps by avoiding an automatic interdependence of pitch and voice timbre.

Not very surprisingly, similar effects seem to hold in instrumental music. In a classical sound example published by Pierce (1983), the psychologist David Wessel demonstrated that the tones in a sequence could be glued together by giving adjacent tones similar timbres. If, on the other hand, adjacent tones differed in timbre, the melodic line broke at several places split into several lines. This effect seems to demonstrate how important it is for singers and other musicians to acquire a complete control over the timbre; if this goal is not achieved, the melodic lines will tend to break in places where they should not.

**HOW MUCH?**

The extent to which a rule affects a performance can be called the *quantity* of the rule. It decides how many *ms* that should be added to or subtracted from a tone's duration, how many *dB* should be added or subtracted from its sound level, etc. What quantity is musically convincing? We have tried to find an answer to this question in two experiments (Sundberg & al., 1991).

In one experiment, a computer program was used that allowed a variation of the magnitude of the effect induced by the application of a rule. The tool was a slider on the computer screen that could be operated by means of the computer mouse. In one extreme position the effect was huge. Then we asked five professional musicians to adjust the slider for each of six rules such that, according to their musical judgement, the effect of the rule appeared appropriate. For all rules, it was possible to select zero as the preferred quantity. Fig. 4 shows the average settings and 95% confidence intervals obtained for five of the rules.

The averaged preferred quantities were mostly close to the default values that we had arrived at when adjusting the complete performance program. More importantly, however, this result indicated that our musicians found that these rules im-
proved the musical quality of the performance; the musicians did not find that a zero quantity of the rule-generated effects that were preferable.

![Graph showing mean preferred quantity Q for the six rules indicated, observed when five professional musicians adjusted Q to what they found was optimum. Each rule was tested on one music excerpt. The bars show 95% confidence intervals. The scale used for the quantity is arbitrary, 1.0 corresponding to the default value for the rule in the performance rule system (Friberg, 1991).](image)

In the other experiment, we attempted to determine the threshold quantities for the effects generated by these rules. Here we presented two versions of the same music excerpts pairwise, as illustrated in Fig. 5. In one of the two versions, one rule had been applied so that an effect of a certain magnitude emerged, and in the other version no rule had been applied. The task of the subject listeners was to determine if the two versions differed or not. The subjects would obviously fail to hear a difference if the quantity was below its threshold value. Each rule was tested in series of contingent pairs, and the first of these presented a very great difference, so as to direct the listener’s attention to the aspect of the performance that was being varied.

SERIES 1

RULE 1: The higher, the higher

![Diagram showing principle used for stimulus presentation in the threshold experiment. Each rule was tested by presenting a series of paired performances of the same music excerpt, one series for each rule. In each pair of versions, the first version represented a deadpan performance (Q=0), while the rule had manipulated the second version to a greater or smaller extent (Q>0). The first pair in each series represented a maximum contrast so as to direct the subjects’ attention to what aspect of the performance was being varied. The figure shows the first three pairs of a series.](image)
Figure 6 presents the results obtained. When the differences were great, none of the musicians thought that the two versions were identical. When the difference was small, most of them failed to hear it, although even when the difference was zero, several thought that they could hear a difference in spite of the fact that there was no difference. Probably, this should be interpreted as a sign that some of the subjects were overly eager to demonstrate a very sensitive musical ear. For most rules, the response curves showed a clear slope down toward zero per cent of “Same” answers for the greatest differences. With the non-musicians, the results were less clear, so that even for great differences many subjects failed to detect any difference.

Fig. 6. Threshold quantities for effects generated by the seven performance rules indicated. The graphs show the percentage of “Same”-answers received in a test where musicians (filled circles) and non-musicians (open circles) compared deadpan and rule-modified versions of the same examples. The abscissa represents the physical difference between the two versions, specified in terms of the quantity parameter Q, which was normalized with respect to the default value used in the program. The curves, derived by processing the data points with a LOGIT program (Bock, 1975), illustrate the overall trend of the data points.

These results appear to suggest that non-musicians need greater effects than musicians for hearing the finer differences between the performances. But it is also possible that musicians and non-musicians hear equally well, although non-musicians
have a harder time analyzing and faithfully verbalizing what they hear. Perhaps they listen as attentively but intuitively, and are incapable of giving a verbal account for what they heard. In any event, musicians are certainly accustomed to listen also more analytically and, of course, they are accustomed to speak about performances.

Let us now digress somewhat and consider a particular practice in musical performance. "Alte Spielweise" is a style of organ playing that has been adopted by some organ players. It is considered to do justice to organ music composed during the 17th and 18th centuries. One of the Swedish advocates has published phonograms where she demonstrates this playing style. On the envelope she writes: "How did one mark the rhythm on organ and harpsichord? It was by stretching those tones on which the beats fall. How much? Well, this of course is a complex question. It depends on the Baroque tone characteristics."

In motorical pieces of music, i.e., pieces with marked sequences of equally short notes, lengthenings of about 20 ms seem sufficient for evoking an effect. Measurements have shown that lengthenings and shortenings of this magnitude are common in music performances (see, e.g., Palmer, 1988). However, such changes of tone duration are difficult to identify as such. Rather, they have the effect of eliminating the mechanical character of a deadpan performance. To be recognized as a lengthening or shortening of a specific note, the perturbation has to be much greater, probably on the order of 50 ms. Interestingly, effects of some 40 to 70 ms were found in a phonogram recording of organ playing according to the "Alte Spielweise". Thus, in this particular performance practice, the musician seemed to prefer to work with clearly recognizable lengthenings and shortenings, even though effects reduced by a factor of 2 or 3 would be sufficient for eliciting musically useful effects.

It is not incumbent on science to tell artists how to make art or to decide what is beautiful and what is ugly. On the other hand, there is no problem for science to tell what the smallest noticeable difference is. It is quite clear that this organist uses effects that are much greater than the smallest perceptible effects. It seems that in this case, not only perceptible but also correctly analyzable effects are strived for. It also seems clear that much finer effects would have been sufficient if the only purpose was to glue tones together that belong together.

**NONSENSE-TEXT SINGING**

When musicians talk to each other about how they think a theme should be performed, they sometimes sing the theme using improvised nonsense texts. The method is also sometimes used by great conductors instructing the orchestra. Let us imagine that a conductor wants to demonstrate how the "Waltz of the Flowers" from Pjotr Tchaikovsky's "Nut Cracker Suite" should be played. Perhaps he would sing the following text:

It seems clear that, in a sense, these syllables are not meaningless. On the contrary, they mean specific things. One way of demonstrating this is to simply change the order of the syllables above and ask if this would be an equally suitable sequence of syllables for the melodic theme:
It seems doubtful that an instruction using this text would be understood in the same way as the previous one. Actually, a conductor using this nonsense text to demonstrate what performance he has in mind would probably be regarded as pretty crazy by musicians. In any event, if these two nonsense texts are not equivalent, they cannot be said to be nonsense.

What do they mean, then? Perhaps, one can see traces of performance rules in them? Before starting to look for an answer to these questions, one thing should be clearly stated. The choice of syllables in improvised nonsense text singing does not carry the entire information. The way in which the syllables are pronounced is another obviously extremely important factor, but here we will not analyze that aspect.

An experiment was carried out in which we asked eleven professional musicians to demonstrate how they thought a set of six music excerpts should be played (Sundberg & Frydén, 1992; pp. 69-84 in this issue of STL-QPSR). We recorded their sung demonstrations on tape and examined their choice of syllables. We found some principles that appeared relevant to the choice.

One principle seemed to be to accompany a transition from one group of tones to another by a change of the syllable. In the tune "Santa Lucia," nine out of eleven musicians changed syllable between the half note terminating the first subphrase and the following second subphrase, as can be seen in Fig. 7. Conversely, very few musicians changed syllable from the first to the second eighth note in the pairs constituting the appoggiaturas, a case of clearly contingent eighth notes. The nasal consonant [m] was often used to mark the end of short tone groups. Fig. 8 shows where this consonant occurred in one of the excerpts, taken from a string quartet by Joseph Haydn. It can be noted that a final [m] seldom appeared on short tones but frequently on tone repetitions. Furthermore, it did not appear between tones that belong strongly together such as appoggiatura tones, i.e., when a non-chord note appears on this stressed beat and is followed by a chord note one scale step down.

There are certainly a number of other observations that also can be made from this material. However, for our present purpose, the material presented so far seems sufficient. It shows that musicians in spontaneous singing of music themes use nonsense texts that are not at all meaningless. The choice of syllable is used to mark the grouping of tones. This result was not unexpected but provides another striking example of the fact that marking of group boundaries is extremely important in music performance.

DISCUSSION AND CONCLUSIONS

This review has shown that grouping is marked in many different ways in music performance. One exception is the rule Faster Upright, which shortens tones initiating an ascending interval and thus moves them closer in time. The rule increases the tempo in ascending scales or triad sequences. The rule Micropauses in Leaps promotes the association of tones that are close to each other in pitch. In cases where the notated bar is a unit of relevance to the musical structure, a singer may mark the first beat in the bars by giving them a small accent produced by a sudden increase-decrease of subglottal pressure. Phrase endings can be marked by lengthening of the final note, while subphrase endings often are marked by a micropause.
Fig. 7. Syllables, transcribed to phonetic symbols, used by the eleven subjects for the various tones in syllabling Excerpt No. 1, Santa Lucia.
An interesting aspect is how this information on grouping is communicated, i.e., what the acoustic code is and from where it may originate. It is obvious that in many cases the code used in music performance is similar to or even identical to that used in speech. For instance, lengthening of the final element in a structural constituent appears also in speech as final lengthening, as mentioned.

The code used in speech is not universal. For instance, termination is signaled by other means than lengthening in some languages. Musicianship, on the other hand, seems largely independent of the player’s mother tongue. This suggests that speech is not the origin of the code used in music performance. The code must originate in another area.

One such area may be locomotion. The anatomy of the final ritard, typically used in performance of motoric music, was studied in a previous investigation (Sundberg & Verrillo, 1980). An average ritard curve was established, normalized with regard to the initial tempo and the duration of the ritard. Later, Kronman & Sundberg (1987) showed that this average ritard curve was almost identical with the curve showing the decrease of footstep rate that characterizes a perfectly preplanned stopping of running. This suggests that the usefulness of the ritard for announcing the approaching end of the piece builds on the fact that the pattern can be recognized as the one which we produce when we stop running. In this sense, the final ritard seems to allude to the listener’s experience of locomotion. To use Clarke’s (1987) terminology, the acoustic code used in the final ritard is iconic.

Physiology of the receiving system seems to explain the code in some cases. For instance, the reason why emphasis is marked by increased loudness and duration in both speech and music performance may be that the perceiving system needs more time to process unexpected than expected information. Likewise, we may speculate that the final lengthening serves the purpose of allowing the brain a bit extra time for processing the completed structural unit.

There would be other acoustic codes used in music performance to convey information about which tones belong together, and where the boundaries are between structural constituents. The four classical Gestalt laws Proximity, Similarity, Good Continuation, and Common Fate would exert their power also in music performance. While Good Continuation seems to be the composer’s headache rather than the performer’s, several cases of Gestalt law can be discerned in music performance. A few examples may be mentioned.

Proximity in time results from the shortening of an upbeat. The Similarity law can explain why discontinuity can be used for marking boundaries and continuity for marking togetherness. The Common Fate law seems to explain the effect that all notes participating in a crescendo sound as belonging together. Is it surprising that the code used in music performance seems to obey the Gestalt laws? These laws would describe how our perceptive system typically processes sensory stimuli, and it would indeed be surprising if music performance could offer an exception to the applicability of these laws.

Summarizing, there are many striking similarities between the acoustic codes used in speech prosody and music performance. The code is sometimes iconic, such as the final ritard apparently alluding to the stopping of running. Sometimes the code seems to simply answer demands raised by the perceiving system, e.g., allowing more processing time for a completed phrase or for a remarkable piece of information. In some cases, the code seems to simply draw on the Gestalt laws.
Fig. 8. Occurrence of various phenomena in eight musicians' syllabing of Excerpt No. 2, the theme of the first Kyrie from J.S. Bach's b-minor Mass: repetitions of the syllable from one tone to the next, with final nasal consonant, with diphthong, with initial vowel, with initial voiced stop consonant, and with initial unvoiced stop consonant.
The shortenings and lengthenings required in order to produce a noticeable effect are quite small. Only 10 or 20 ms seem to be enough, at least for musicians. However, in order to realize exactly what happens in a physical sense, i.e., which tones were lengthened and which were shortened, it seems that the effects need to be two or three times as large, i.e., on the order of magnitude of 40 ms. It seems that some organists playing according to the ideal of the "Alte Spielweise" use effects of such magnitudes.

Although the effects that other professional musicians seem to prefer are small, the marking of structure seems exceedingly important. When musicians show how they want a theme performed, they mark the structure by means of the choice of syllables: change of syllable is one out of several ways in which group boundaries can be signaled. Another alternative is to insert a syllable ending on a nasal consonant, e.g., [pad].

From these observations we can conclude that music listeners seem to expect some gentle help from the musicians in the cognitive work needed to realize, how the piece is constructed. This help is given in the form of small hints showing where the boundaries are between different structural constituents.

This marking of structural boundaries is not unique to music. The same happens in architecture, where boundaries between structural constituents are often marked. For instance, the boundaries between windows and walls often are marked by means of colours; window frames are generally painted in a colour different from the adjacent wall. Speech is another example; the importance of marking of structure is a well-known fact in speech science. It is considered a fundamental ingredient in what is generally referred to as prosody. Two examples were already mentioned: micropause for a comma and final lengthening for a period. Literature is another example where structure is carefully marked. An extra space is inserted between the words, a dot after each sentence, and each paragraph is started on a new line and each chapter is started on a new page.

It seems that the marking of structure is a commonplace phenomenon in human communication systems. Still, it may appear as slightly remarkable that it is considered so important because even without such signs, we tend to realize where the boundaries are. This, precisely, may be the reason why we get so agonized when we hear a poor musician, who reveals from his way of playing that he does not understand what belongs together structurally and what does not, or when we hear someone reading a poem aloud in such a way that it is obvious, that the reader does not understand what he/she is reading. The reasons for such reactions is a question for psychology. It may be related to the strong emotional impact that music has on its listeners. For the moment, we can conclude that gluing tones together that structurally belong together, and marking the boundaries between tones that do not, seems to be an important part of musical performance.

ACKNOWLEDGMENTS

This paper was originally presented at a symposium at KTH in May 1991, arranged by the Music Acoustics Committee of the Royal Swedish Academy of Music. Carol Krumhansl kindly provided helpful comments on the manuscript.
REFERENCES


