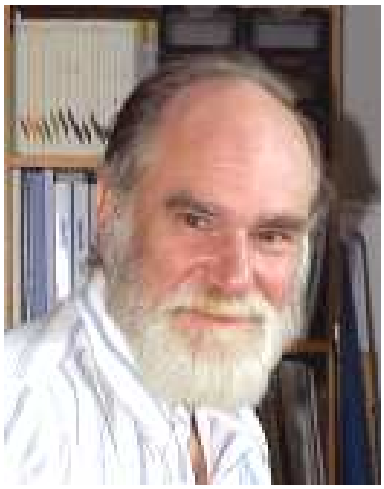
MUSIC ACOUSTICS



Johan Sundberg
Professor of
Music Acoustics

Overview

The research within the music acoustics group runs in three main streams: solo and choral singing, stringed instruments, and music performance. While running in parallel for periods, they often merge and interact. An experimental approach has long been a signature of the research. In particular, experimental work with "analysis-by-synthesis" has proved to be useful.

In the area of music acoustics, the trend continues toward more sophisticated synthesis strategies, based on a direct mathematical description of the instrument/voice. This strategy commonly referred to as numerical or "physical modelling," would be generally adopted in the future at most major synthesiser laboratories. We have started physical modelling of the singing voice in terms of articulatory synthesis, controlled by physiologic control parameters such as jaw opening and tongue shape.

The interest in numerical modelling has revealed a lack of reliable and detailed experimental data in many areas. Our research group is one of very few in the world that works specifically in the area of basic research on musical instruments and sound. Hence, basic research is an important responsibility of our group. Examples are the function of control systems used in actual performances, such as breathing behaviour in singers, or bow movement in the playing of string instruments. Lack of basic knowledge of such control systems forces the experimenter to explore an overwhelmingly rich parameter space, and the chance of real success in reasonable time seems limited.

Our research on musical performance by computer modelling is a long-term project. Its focus is presently to compare characteristics of locomotion in running with timing in music performance. Our generative music performance grammar and music examples illustrating the effects on synthesised performances is accessible on the Internet.

Our overall aim is to continue the established experimental tradition in the study of the voice and the stringed instruments (now needed more than ever, it seems), while striving to take part in the development of the

new synthesis methods. As before, the parallel work on musical performance bridges several of these tasks.

Voice

Breathing and phonation

Breathing and phonation is one of the core themes of the voice research in our group.. Lung volumes used by professional classical singers have been measured under realistic performance conditions. Monica Thomasson has analysed consistency of exhalatory and inhalatory breathing patterns in professional operatic singers, using inductive plethysmography to record rib cage and abdominal wall movements. The results showed a high degree of reproducibility, when the singers repeated the same phrase. Also, while some singers rely mainly on rib cage movements to increase lung volume, others consistently use a combination of rib cage and abdominal wall movements.

Jenny Iwarsson's doctoral project concerns the relationship between breathing behaviour and phonation in pathological voices. As the frequency of inhalations seems a relevant factor, she investigated how accurately inhalations could be heard from audio recordings of female subjects. The material consisted of six recordings of normal voices and six recordings of patients with vocal nodules. The result showed that a listening panel of 30 logoped students managed to hear more than 90% of the inhalations. Thus, listening to a recording of a patient's voice seems sufficient for determining whether or not the patient has a normal frequency of inhalations. This was the sixth and final part of Iwarsson's doctoral dissertation which she defended January 19, 2001.

Timbral ugliness

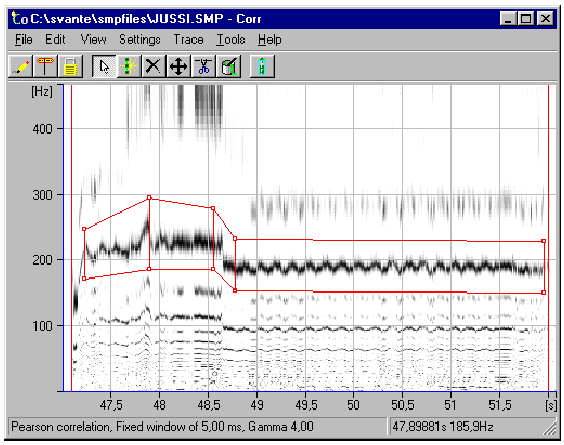
Timbral beauty is an important component in vocal art. Some singers possesses it to extreme extents and other lack it almost entirely. Johan Sundberg compared two recordings of the same excerpt, one by a successful singer and one by an amateur singer. A panel rated the timbral beauty of the voices on a 10 cm visual analogue scale and judged the professional singer's timbral beauty as 7 and the other singer as 1, implying an almost extreme ugliness. Acoustic analysis revealed that the ugly voice lacked a singer's formant and had a weak fundamental, presumably due to hyperfunctional phonation. Also his vibrato was irregular and his mean fundamental frequency deviated occasionally from the tuning of the accompaniment. The singer's formant helps the singer to be heard when accompaniment is loud. Hyperfunctional phonation suggests that phonation is difficult. An irregular vibrato causes the perceived pitch to be unstable. Thus, these differences between a beautiful and an ugly voice suggest that functionality is relevant to vocal aesthetics.

High-speed imaging

High speed imaging produces thousands of frames per second, thus causing difficulties in extracting relevant information. Typically, a minute part of the image is used, such as the position of a section of the vocal folds, while all other information is disregarded. Svante Granqvist has written a program that analyses the complete image from different points of view. Thus the vibration frequencies of all glottal structures can be documented, or the vibration phase relationship between different structures. The method has been tested on normal and dysphonic voices. Among other things, the results show out-of-phase vibrations in glottal structures and that vibrations in tissues adjacent to the vocal folds.

Hunting Fo

Tracking the fundamental frequency is a classical problem in acoustic analysis of speech and music. The problem is particularly impossible to solve when a voice is aperiodic. Svante Granqvist has developed a program that provides a valuable tool in dealing with such problems. The idea is to correlate the signal contained in one time window with that contained in another time window. The result is shown as a *correlogram* with time on the x-axis, frequency on the y-axis and correlation in grey scale on the z-axis. The fundamental frequency of a vocal signal thus appears as a black trace. In cases of aperiodic voices, the correlation trace is widened and pale. The method has been tested on synthesized voices and on real voices with different types of pathology.



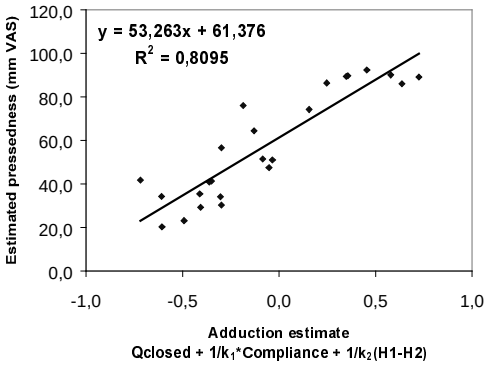
*Correlogram of a section of a recording of the late Swedish tenor Jussi Björling singing without accompaniment. Note that in the final tone the second partial is close to the first formant, which produces a ghost trace at F0*1.5.*

Singing styles

The properties of the human voice source is another core theme in our voice research. It can be assumed to differ depending on the vocal style. Margareta Thalén, professional singer and singing teacher, has studied her voice source in classical, pop, jazz and blues styles of singing as compared to her phonation characteristics in pressed, neutral, flow and breathy phonation.

To check the representativity of the recorded material, an expert panel classified her voice samples in terms of the intended styles. The results showed the expected results for classical and blues styles while pop and jazz styles were sometimes confused. The listening panel also estimated the degree of phonatory press/hyperfunction.

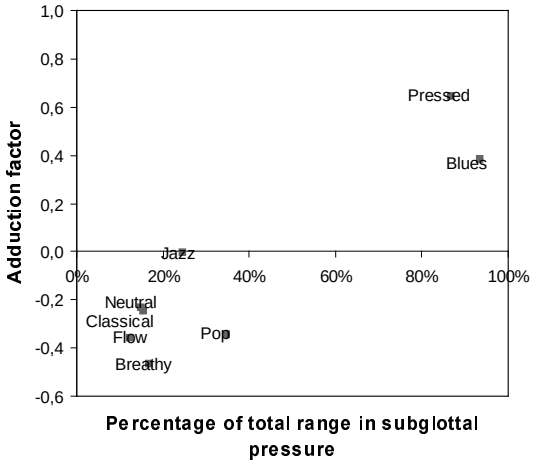
The voice source was derived from the flow signal by means of inverse filtering and the underlying subglottal pressure was estimated from the oral pressure during the occlusion for the consonant /p/. Closed quotient, glottal compliance, i.e, the ratio between the air volume displaced in a pulse, and the level difference between the first and the second partial of the source spectrum were measured. The results were expressed as ratios of the entire variation range of these parameters at each pitch and degrees of vocal loudness. Then the results were averaged across pitch and loudness. As illustrated in the figure, classical



Relationship between mean ratings of phonatory press and a combination of three voice source characteristics measured from inverse filtering.

was similar to flow phonation, pop and jazz to neutral, and blues to pressed phonation.

The glottogram parameters related to the glottal adduction, i.e., closed quotient, compliance and dominance of the fundamental, were compared with the ratings of phonatory press. A combination of these three glottogram parameters yielded a correlation of 0.900. This promising result suggest the need for making similar comparisons for other voices.



Phonatory characteristics of the indicated styles of singing and the indicated phonation modes. Subglottal pressure is expressed as percentage of the total range of variation observed averaged over pitch and loudness conditions. The adduction factor is a weighted sum of closed quotient, glottal compliance and the level difference between partials #1 and 2 of the voice source spectrum. It can be seen that with regard to subglottal pressure the blues style is closed to pressed phonation while classical is similar to flow phonation.

Voice of pre-school teachers

Pre-school teachers develop voice problems more frequently than most other profession groups. The reason can be assumed to be associated with their acoustic environment. Svante Granqvist has adopted Sten Ternström's method to record a voice in noisy surroundings. The speaker carries two microphones, one in each ear. Adding the left and right signals enhances the speaker's voice, while subtraction cancels out the speaker's voice, leaving only the environment sound. Using this method Granqvist estimated the speaker's mean F0 and overall phonation time. These data are needed for assessing the vocal demands encountered in this group of speakers.

Voice therapy without voice therapy

The voice is affected by several non-vocal factors. Dr Anke Nienkerke Springer, University of Dortmund made her doctoral dissertation on nine children, age 5-7 years, who had a deviant voice according to their parents and/or school teachers. The therapy focussed on the psychological family conditions. She offered 28 family therapy sessions of 60 min each to the parents and recorded the voice of the children before and after the therapy period. Johan Sundberg and Dr Anita McAllister, Huddinge University Hospital, analysed the voice samples acoustically and perceptually and compared the results before and after therapy. Significant improvement was found in several parameters, such as hoarseness, phonatory press, gratings, and stability. Changes were found also in some acoustic parameters. The investigation was presented at the Annual Symposium Care of the Professional Voice in Philadelphia in June where it was awarded the "Best Poster" price.

Another example of the relevance of non-vocal factors to vocal function was documented by two naprapat students who made their thesis work under Sten Ternström's supervision. They recorded the voice before and after a 30 min session of massaging the back and neck of 15 subjects. A matching control group rested on a sofa during the same amount of time. Measurements showed no voice changes in the control group, while the experiment group showed two significant changes, a lowering of the mean speaking F0 by about 1 semitone and a lowering of the mean speaking sound level by about 1 dB. A reduction of the speaking level of 1 dB tends to lower the mean F0 by about 0.4 semitones. The observed lowering of mean F0 suggests that the subjects' pitch raising muscles became more relaxed by the massage of the back and neck.

Voice source and subglottal pressure

Subglottal pressure is one of the main control parameters of the glottal voice source. Two investigations were completed regarding the effects of subglottal pressure variation.

Peta White analysed the effect on the overall spectrum slope and found that a 10 dB decrease of the level of the first formant was associated with a 17 dB decrease of the level near 3 kHz.

Two logoped students from the Lund University Hospital, Anja Morell and Ellinor Fahlstedt, measured the effects on the flow glottogram of subglottal pressure variation in 15 female and male untrained voices. The results were similar to those previously found for professional baritone singers although the scatter of the phonatory parameters were greater. The final report of this investigation is under preparation.

Voice Research Centre

The KTH Voice Research Centre is arranging the 4th PanEuropean Voice Conference that will take place in Stockholm, August 23-26, 2001. About 500 requests for the Second Announcement were received, and about 130 abstracts were submitted. The conference will focus on five main themes: (1) "Basic Biological Aspects: Application to Voice Disorders"; (2) "Voice Source: Acoustics and Imaging"; (3) "Assessment & Efficacy of Voice Treatment"; (4) "Singing & Professional Speaking Voice"; and (5) "Non-verbal Communication". Researchers in the doctoral career will present their work in a special session, "Dissertation Pipeline: in & after".

The Centre published *Child Voice*, the proceedings of the fourth symposium it arranged at KTH in 1998. This book provide a survey of the present knowledge in this area and was edited by Peta White. The book can be ordered from the KTH Voice Research Centre.

Instruments

Arching of the violin top plate

The violin top and bottom plates are arched, although the form of the arching varies between instruments. These differences are significant to the quality of the instrument, affecting several basic propertis. The arching of the top plate increases the strength so that it can support the great force excerted by the strings for several centuries. Also, it affects some of the violin resonances that are decisive to the quality. The arching has the effect of simultaneously increasing the width and shortening the legnth of the plate during a phase of the vibration cycle.

Using optical measurement techniques the significance of the arching on the vibration properties has been studied by Erik Jansson in cooperation with professor Nils-Erik Molin at the department of Experimental Mechanics at the Luleå Technical University. The results revealed that these effects on the width and length of the plate clearly cooperate with the vibrations perpendicular to the plane of the palte and bottom.

If we imagine that the violin body is cut perpendicularly to its length axis at the bridge, the shape is roughly similar to a tube with an elliptical cross section. During the vibration cycle the shape can be flattened so that it assumes a more elliptical shape, or thickened so that the shape gets more circular. The bridge excites the top plate by a rocking motion, such that the shape alternatly gets more elliptic and circular. A nodal line is created between the bridge feet, where the plate does not move. These unexpected finding will be presented in the doctoral thesis by Anna Runnemalm at the Luleå Technical University.

The bridge hill

A characteristic of good violins is a rather wide resonance in the 2500 – 3000 Hz range. The resonance has generally been assumed to reflect the resonance characteristics of the bridge itself, hence the term ”bridge hill”. New experiments, however, with a Stradivari violin have revealed that the bridge hill is mainly dependent on the properties of the top plate, particularly its arching, while the contribution from the bridge resonance is modest. The shape of the arching determines how many fibers are cut in the plate, when the builder planes it to the intended shape. Possibly, the number of cut fibers is a factor influencing the frequency and shape of the bridge hill.

Violin bows

It is well known that the properties of the bow has a great effect on the timbre of bowed instruments, be it a violin, an alto violin, a cello or a double bass. The physical reasons for this influence has remained an unanswered question.

Anders Askenfelt and professor Knut Guettler, Norway’s Music State Conservatory, Oslo, have found a new promising mechanism that may answer the question. When the bow is moved over the string, a periodically varying frictional force is generated between the bow and the string. This is the force that drives the string and maintains its vibration and hence the tone. The same force also excites the bow itself. Depending on different combinations of frequency and damping of these bow resonances these vibration will be more or less pronounced in different bows. Simulations have shown that vibration perpendicular to the string, i.e., vibrations that affect the force excerted by the bow hair on the string that are most likely to affect the string vibrations and thus also the tone quality. Even a relatively moderate periodical variation of this bow force may be sufficient to affect the relationship between that portion of the vibration cycle where the string sticks to the bow and that portion where the string slips back. Such a modulation of the relation between the slip and stick phases will be reflected directly in a time dependent variation of the amplitudes of certain partials in the sound produced. Admittedly, the effect would be small, but so are the measurable differences between a top quality and a mediocre bow. Hopefully, these preliminary observations can be corroborated experimentally in a well controlled analysis of a representative selection of bows of differing quality. Listening tests with manipulated violin tones should be carried out to determine how small modulation of the bow force can be perceived, and also to determine what the optimum is.

Music performance

Legato and staccato in piano

Roberto Bresin has studied the performance of *legato* and *staccato* articulation in professional player's recordings on a Yamaha Disklavier. The recordings were made at the Music Conservatory *Benedetto Marcello* in Venice and on a Bösendorfer grand piano in Vienna.

Computer analyses showed quite systematic effects. Thus, *staccato* tones were realised by lifting the key at a certain percentage of the inter-onset time, regardless of tempo. The moment at which the key was lifted varied depending on context. In a sequence of staccato tones the key was lifted after 25% of the duration of the tone. In isolated staccato tones the key was lifted earlier and in an ascending series of pitches they were lifted later than in a descending melodical contour.

Emotional expressivity

The Director Musices program is a generative grammar of music performance, created by Anders Friberg. It automatically converts music files into sounding performances possessing musical expressivity. The reason is that the program introduces a number of deviations from the score that a performing musician will typically make. These deviations that seem required to make the performance expressive.

Roberto Bresin and Anders Friberg have implemented in the Director Musices grammar Juslin's and Gabrielsson's qualitative observations regarding emotional coloring of music performances. The result is an emotional coloring of the performance that can be correctly identified by listeners. Thus, the grammar generates performances of a given piece that sound happy, sad, solemn, etc. One major factor corresponds to tempo and loudness. All remaining factors can be derived from the musical context as defined by the score.

Bresin defended his doctoral dissertation Virtual Virtuosity on December 1. Roger Dannenberg, Carnegie Mellon University, acted as Faculty Opponent.

Tempo drift in drumming

In her investigation of timing in drum performance, Sofia Dahl observed that drummers typically produce small changes of the tempo over time. Together with Monica Thomasson and Svante Granqvist she carried out an investigation on the perceptibility if such tempo changes. Series of click sequences were presented at a starting rate of 120 beats per minute to ten musicians and ten naïve music listeners. Different amounts of tempo changes occurred in the sequences, the length of which was varied by a computer program. When the sequence was about 5 sec long, listeners could detect tempo changes as small as 0.2% per interval. Some listeners even managed to detect changes of 0.01% per interval if the sequence was about 20 s.

Estonan folkloristic singing

Jaan Ross, professor of musicology at the Tartu University, Estonia, and Anders Friberg have compared the rules in the Director Musices generative grammar of music performance with performances of Estonian (Finnish-Baltic) folk songs. These songs are sometimes called Kalevala-songs. The oldest ones have been estimated to be between two and three thousand years old. This tradition has developed rather independently of the European classical tradition. Hence, it was interesting to find out to what extent the performance of these songs showed any similarities with those generated by the Director Musices grammar, which mainly reflects European classical performance tradition**.**

The songs in this repertoire all have a similar structure. This implies that lyrics and melodies can be combined rather freely. This offers a unique possibility to study the interrelations between text and melody with respect to the performance. Four of the performance rules contained in the Director-Musices program seemed potentially relevant for the songs: Inegales, Melodic charge, Faster Uphill, and Phrase. Three new rules were defined reflecting articulation, Estonian prosody and performance practice in the songs. A covariance analysis showed that these seven variables explained between 20 and 50% of the variation in tone duration. Of the Director-Musices rules, only Inegales contributed significantly to the variance. It can be concluded that the durational variation in these songs mainly depends on linguistic factors while the influence of musical factor, such as phrasing and direction of the melodical movement, is relatively small.

Violin vibrato

The vibrato is an important expressive parameter in violin playing, and violinists have rather large possibilities to vary vibrato parameters. The guest researcher Julieta Gleiser, Argentina, and Anders Friberg, analyzed vibrato rate and extent in violin performances by four professional violinists who all played their own version of six pieces. Using Svante Granqvist's Correlogram program the continuous variation of vibrato rate and extent for each tone. The analysis indicated that the violinists varied both rate and extent within wide limits, though often with a strategy that differed between players. The individual variation was smaller for long than for short tones. The extent seemed partly related to loudness, such that loud tones were played with a wider extent. The vibrato tail, previously observed by Eric Prame, i.e., an increase of the rate toward tone endings, was observed also in their material, particularly in some of the players.

Radio Baton

Professor Max Mathews, Stanford University, has developed a sequencer, the *Radio Baton*. It is controlled by two sticks that are beaten against a plate. The positions of the sticks above the plate are sensed in three dimensions. The Radio Baton thus allows control of a performance in a manner somewhat similar to that of a conductor. One stick gives the beats while the other controls the dynamics. Gerald Bennett, professor of composition at the Zürich Music Conservatory, Max Mathews, Johan Sundberg and Anders Friberg met during a week in Zürich to combine the Radio Baton with the Director Musices performance grammar. The combination seems to have a high potential value for music pedagogy, where teachers often have to split their attention between following the student's playing and playing the accompaniment part in the piano. The work will be continued during 2001.

GERM

A complete model for musical performance has been developed by music psychologist Patrik Juslin, Uppsala University. The model, called GERM, contains four basic components: **G**enerative rules reflecting the musical structure, 2. **E**motional rules, 3. **R**andom variations caused partly by variations in the player's internal clock, and partly by variations associated with the relation between neural impulses and the corresponding muscle contraction, and 4. **M**otion related rules. Anders Friberg and Roberto Bresin have cooperated with Patrik Juslin to build a complete rule system that implements the GERM-model. Most efforts have been spent on the random factor, which is a rather untouched area. A preliminary evaluation in terms of a listening test was carried out.

Temporal aspects of perception

Eric Prame has continued his investigations of temporal aspects of perception by studying the perception of 100% AM of broad band noise and of AM and FM of complex tones and their dependencies on the modulation frequency. This has been done both for modulation frequencies above 8 Hz, the region of "roughness", as well as below that frequency. Previous studies of the roughness region has shown many similarities between AM and FM, which, however, is not the case in the region below 8 Hz. The phenomena in the lower region are also less well known. At 4 Hz, for instance, the perception of the increasing phase of an AM modulation period is clearly dominating. More astonishing is that some asymmetry also can be found in FM. This observation is important for understanding the vibrato phenomenon. One explanation is neurophysiological. The dividing frequency of about 8 Hz between the two regions is based upon the research of professor Robert Efron, who discovered that the minimum duration of a perception is about 130 ms; when inter-onset intervals between repetitive stimuli are clearly shorter than 130 ms, perception of roughness emerges. Robert Efron visited Eric Prame during one week in the summer to discuss in more detail his research.