Appeared in: (2000). Psychology of Music, **28**, 23–42.

A New Method for Analysing and Representing Singing

Stefanie Stadler Elmer, Institute of Education, University of Zürich &
Franz-Josef Elmer, Institute of Physics, University of Basel

Abstract

In psychological and cross-cultural (*e.g.* ethnomusicological) research the analysis of song-singing had always been an intricate and serious obstacle. Singing is a transient and mostly unstable patterning of vocal sounds that is organized by applying more or less linguistic and musical rules. Traditionally, a sung performance has been analyzed by mere listening and by using the western musical notation for representing its structure. Since this method neglects any in-between categories with respect to pitch and time, it proves to be culturally biased. However, acoustic measures as used in speech analysis have had limited application and were primarily used to quantify isolated parameters of sung performances.

For analysing and representing the organisation of pitch in relation to the syllables of the lyrics, and its temporal structure, we devised a computer aided method in combination with a new symbolic representation. The computer programm provides detailed acoustic measures on pitch and time. We reduce the redundancy of the detailed information by a notation system that shows pitch and time each on a continuous scale, including glissandi, breathing, joint singing, and instructional help. By combining acoustic with auditory analyses, this method allows to describe reliably sung performance's structures with respect to the organisation of pitches, together with syllables, and their timing. The resulting configuration of data includes qualitative aspects such as stable and unstable pitches. Such microanalytic descriptions are very useful for studying the nature of sung performances, their structures, and processes of change due to learning and development.

Key words: vocal sound analysis, singing, microanalysis, structural approach, musical learning, cultural psychology, cross-cultural research, ethnomusicology, methods

Introduction

Partly due to the renaissance of Vygotsky's work, there is a growing interest in cultural issues and in socio-cultural processes that promote individuals' acquisition and internalisation of cultural products and rules. One such product that exists in every human culture is the social tradition of song-singing that is mediated and transformed from one generation to the next. From early on in life, children familiarize with the musical conventions, especially with the child-directed practice of children's songs (*e.g.* lullabies). At amazingly early ages children produce vocal musical sounds such as recognisable songs and song fragments. For instance, children's productions of the traditional German song 'Hopp, hopp, hopp, Pferdchen lauf Galopp' are reported by Stern (1914/65) about Guenter at age 1;10 yrs., by M. & H. Papoušek (1981) about Tanja at age 1;1 yrs., and by Stadler Elmer (1997) about Ursina at age 1;8 yrs. These children performed vocal sounds that were structured in such a way that the intended song was easy to identify. Of course, these early products were not yet properly organized with respect to all of the complex and hierarchical rules that make up a song. The learning of a culture's song-singing rules is a long-lasting process of enculturation that basically includes linguistic (lyrics) and musical (melody) rules, both combined by meter (pulses that are periodically stress and unstressed), and social rules that regulate participation by jointly matching and synchronising vocal pitches.

In children the vocal structures are in progress towards integrating these rules. When stimulated to participate in recurrent and joyful singing rituals, young children readily adapt their vocal structures to these joint activities. Usually, the adult guide is able to recognise the newly acquired song fragments in the child's vocalisation as a response to their communication and shared musical experiences. Moreover, the adult's interpretation and understanding of the child's singing is based on cultural concepts or knowledge. With this situated and cultural knowledge, a minimum of cues is sufficient in identifying the child's intended (musical and linguistic) meaning of her or his vocalisation.

Usually, there is no need to reflect on this functional tendency for overinterpreting precultural behavior. On the contrary, interpretation by cultural categories or concepts helps reducing the wealth of perceptually available information, allows distinctions among objects and events, and adds meaning to them. In this sense, musical training aims at gaining and refining categorical perception and musical concepts within a given cultural system. It is a collectively shared semiotic system, and akin to language, it provides the very basis for communication and the social practice of sound-making.

Siegel & Siegel (1977b) showed that highly trained musicians had a strong tendency to rate out-of-tune stimuli as in tune. Their attempts to make fine, within-category judgments were highly inaccurate and unreliable, whereas they differentiated well between musical categories. The musicians were not aware of their biases in categorising stimuli in terms of musical concepts. Since these authors proved that nonmusicians have great difficulty in discriminating even between musical categories (1977a), they conclude that the phenomena of categorical perception to be a result of musical training. It is functionally similar to phonemic categories for speech.

Many developmental and learning theories, especially the ones in the tradition of Piaget (see *e.g.* Beilin & Pufall, 1992; Smith, 1996) favour structural approaches in the sense that actions are analysed as organised units in order to study underlying mental processes and structural changes. According to such theoretical approaches, we consider singing as a complex organised action whose structures are guided mentally (*e.g.* Imberty, 1996; Stadler Elmer, 1998). Whenever we want to describe and explain unconventional singing either by children or by people from other cultures we are faced with a rather paradoxical situation. Psychological investigations into musical issues require musical knowledge. But it is this knowledge that turns out to be an obstacle in understanding children's, novices', and general non-western musical behavior and cognition. Musical training implies strong biases towards culturally established meanings and symbols which may not be present in the phenomena. Seashore (1938) impressingly demonstrated the discrepancies between a singer's interpretation of a score and the listener's judgments of intonation. He concludes that 'It is shockingly evident that the musical ear which hears the tones indicated by the conventional notes is extremely generous and operates in the interpretative mood. ... the matter of hearing pitch is largely a matter of conceptual hearing in terms of conventional intervals, ...' (p. 269). Deviations from expected pitches are tolerated and are a means for aesthetic expression, particularly in artistic singing with vibrato (cf. *e.g.* Seashore, 1938; Sundberg, 1982).

To conclude, the researchers need musical knowledge for understanding musical phenomena, but this knowledge may misrepresent the phenomena due to the highly culturally influenced mind that tends to overestimate cultural meaning which might not be intended by the producer of music.

The problem we discuss in this paper is, therefore, to cope with our musical mind in a way that allows to minimise its high propensity of unreflected cultural interpretation. In addition, any solution whatsoever should meet the requirements of approaching singing as an action with its inherent organisation. First, we discuss the traditional methods for analysing and representing 'musical' products, especially singing, and its shortcomings. Second, we show that computers can be used as excellent research tools to eliminate the researcher's musical mind at an important stage in the research process. Third, we propose to reduce the detailed acoustic data by a notation system that is more differentiated than the conventional one. Then, we illustrate this method with an example of a young child's song-singing. The utility of the method is discussed by comparing it with a professional musician's transcription of the same song-singing.

Beyond general delight of children's singing, scientific interest in this behavior can be traced back to the beginning of this century. Yet, the fleeting and transient nature of singing and the related problems of assessing its intricate nature seem often to have kept researchers from conducting systematic investigations. Our method promises to open new ways of gaining insights into psychological processes in this cultural domain.

Traditional Methods for Analysing Singing

In his review of the ways singing can be assessed, Welch (1994) distinguishes between 'machine-based' and 'human-based' analyses. Welch's term 'machine-based' analyses refers mainly to those methods which assess the underlying physiological bases of singing such as electrolaryngography. By 'human-based' analyses he refers to auditory ratings done by professional musicians about 'goodness of fit' with regard to the rules of the musical culture.

Besides these two methods, researchers want to study *what* children sing and how the structures of their singing change as a matter of progressive adaptation to their socio-cultural surrounding through learning and development (see *e.g.* overviews by Anderson, 1991; Atterbury, 1984; Stadler Elmer, 1996). However, the nature of such enculturation processes is a key interest in the fields of development, education, and cultural learning in psychology. In a cross-cultural and ethnomusicological context, it would be important to have research tools that permit some kind of culture-free and reliable access to the actual musical productions.

Heinz Werner (1917), in his pioneer work on children's singing, already had to deal with the following two problems: Analysing a form of singing that does not yet adhere to adult conventions, and of representing the sounds' transcription on paper. He analysed recorded singing by repeated listening, and he used quarter pitch notes for representing inaccurate singing. It appears that most researchers since have pursued the same procedure. That is, if possible, they record and then auditorally analyse the singing, using conventional musical notation with some additional symbols such as quarter pitch notes and a cross sign for speechlike reproductions (*e.g.* Moog, 1967, 1968; McKernon, 1979; Davidson, McKernon & Gardner, 1981; Davidson, 1994; Davies, 1986, 1992; Kelley & Sutton-Smith, 1987). This method, thus, approaches song-singing on the basis of purely auditory analysis. Sometimes, assistance from professional musicians is mentioned (*e.g.* Davidson, 1985), and they used musical instruments as a means of comparing and improving conceptualisation. Nevertheless, this procedure does not prevent from perceiving categorically. Somewhat surprisingly, automatic tuners are only rarely used (Flowers & Dunne-Sousa, 1990; Stadler Elmer, 1990). Such apparata are made for tuning musical instruments. They receive sounds, analyse them, and show the results immediately in form of a light flashing on a temperate half-tone keyboard bar. Although it is a useful supplement to a purely auditory analysis, it is tedious because it only reacts to immediate input and is able to handle only broad pitch categories.

Papoušek & Papoušek (1981) tackled the problem of analysing early pre-musical and prosodic vocalisation with detailed and extensive procedures. They applied several acoustic methods that are widespread in speech analysis, and combined them with auditory analysis represented by musical transcriptions. By that, the resulting representations show a wealth of information about a sung performance's structure. Papoušek & Papoušek use this combined method of auditory and acoustic analyses not only for explicit singing contexts with children but also for highlighting musical elements in preverbal communication (*e.g.* M. Papoušek, 1996). In fact, their solution for analysing and representing singing is closest to what we propose in this paper.

There are other ways computer-assisted tools are used in the context of singing analysis: In order to specify 'good' from less competent singers, Sergeant (1994) proposed some acoustic features that he gained with the aid of a computer program (devised by D. Howard) that extracts the fundamental frequency of the sung response. This program seems to work similar to the one presented below. However, the author was interested in analysing acoustic features of children's single pitch matching and in comparing trained with untrained voices at the basis of single sounds produced. He was neither interested in analysing larger vocal units such as song-singing nor in working out a notation system for such larger units. The same is true for research that focuses only selected acoustic features but not singing as entire musical event with regard to vocal pitch and its timing. E.g. Trehub, Unyk, Kamenetsky, Hill, Trainor, Henderson, & Saraza (1997) were interested in quantifying the average pitch level at which parents produced infant-directed songs. They asked a musicologist to identify the tonic or principal pitch (of the song's key) which was then measured by computer tools, yielding the fundamental frequency. They focused a single parameter and were neither interested in the organisation of pitches and time, nor in the stability of the tonic's fundamental frequency, *i.e.* changes of the musical key, throughout a performance.

Limitations of the Traditional Methods

Analytic methods that are based on purely auditory analysis have two major limitations. The first and most restrictive one lies in the aforementioned analyser's propensity to preconceptualise and categorise unconventional singing into musical concepts. Since we know that children, novices, and people with nonwestern education do not have full perceptual and conceptual command of western musical concepts and of rules in the same way as trained individuals habitually do, a description of their vocal expressions should take this fact into account. Auditory analyses and transcriptions solely based on musical concepts presumably are biased towards overestimating the cultural concepts and rules. They do not adequately capture the relevant developmental or learning phenomena.

The second limitation is related to the first one. Since conventional music notation allows only that kind of singing to be represented which already fits somehow into this preset frame of symbols and categories, all singing that is outside what we traditionally consider by a narrow interpretation of this concept or that has not yet been sufficiently adapted to the cultural forms of expression, such as poor, out-of-tune, or nonwestern-styled singing, cannot be assessed by these methods. So far, this problem has been tackled by verbal descriptions of children's poor or not so well-developed song-singing, or by employing descriptions using roughly ascending and descending lines or some additional symbols within the frame of the traditional musical notation system.

There is a long tradition of using computer-assisted methods for speech analysis (*e.g.* Hess, 1983; Tohkura, Vatikiotis-Bateson, & Sagisaka, 1992), especially with regard to acoustic features of intonation (speech melody or prosody) (*e.g.* Helfrich, 1985; t'Hart, Collier, & Cohen, 1990). These methods were applied, for instance, to substantiate acoustically prosodic features of adult's infant-directed speech ('motherese') (Fernald & Simon, 1977; Garnica, 1977) and became influential in this domain (*e.g.* Fernald & Simon, 1984; Fernald & Kuhl, 1987; Fernald, Taeschner, Dunn, Papoušek, de Boysson-Bardies, & Fukui, 1989). Further applications concern prosodic analysis of infant-parent dialogues (*e.g.* Papoušek & Papoušek, 1989; M. Papoušek, 1994, 1995) and developmental aspects of children's speech (*e.g.* Smith & Kenny, 1998).

Whereas computer-assisted methods for analysing acoustic features of speech melody are widespread, they only rarely are applied systematically to analyse and represent sung performances. When applied to singing, they tend to evaluate only single and isolated parameters (*e.g.* Trehub et al., 1997). Moreover, although singing and speaking share some common features, they differ markedly in the conventions on how to organise temporally the pitches together with the syllables of the lyrics. Methods for speech analysis have to be adapted to the specific nature of singing. For example, the analysis of singing has to account for the fact that each syllable can have a different average pitch level (which is less important in prosody), and that the syllables are more or less metrically timed. The variations of pitch within a syllable is an important feature of prosody, but it is ignored in traditional methods of singing analysis. In order to account for different patterns of pitch variation, we use different symbols for assessing the sung pitch quality (see section , Tab. 1, and Fig. 2).

As an exception, Papoušek & Papoušek (*e.g.* 1981) and M. Papoušek (*e.g.* 1981, 1996) examined vocalisations (infant-directed speech and singing, young children's singing) with regard to several structural aspects. They combined acoustic with auditory analyses for the same vocalised unit. However, they used acoustic measures, but represented the musically relevant features (pitch and time) with conventional music notation. As discussed above, for certain research purposes we consider musical notation not appropriate for describing actually performed song-singing.

In doing research that focuses on children's and novices' singing and its process of enculturation, we are, thus, faced with the following epistemological and methodological problem: *How can we analyse and represent adequately the organisation of pitch and time in pre-conventional singing since it does not yet fit into the frame of our musical notation system and conceptual conventions connected with it?*

Similarly, this question applies to research carried out in cross-cultural contexts involving musical sound patterns that do not follow the habitually used musical concepts and rules.

Our aim was, thus, to devise a method that would allow not only more objective and reliable measures with computer-assisted analyses of sung performances, but also to develop a system for representing the organisation of the musically relevant features by more differentiated symbols than the culturally given musical notation system. This method should provide reliable descriptions of a sung performance's constituent components, foremost the configuration of pitches, their timing, together with the sung syllables or lyrics.

The Advantage of Computers: The Elimination of the Musical Mind

During the twentieth century, important progress was achieved in developing and improving the technical recordings of sounds, and for the last few decades computers serve as excellent research tools in approaching high quality of acoustical analysis of sounds. These technical possibilities permit and support a de-conceptualisation of our culturally coined perception and conception of complex sound phenomena such as human speech and singing.

To be clear, the problem as we see it, is not a matter of subjectivity in conceptualising singing. Two trained musicians may easily come up with high agreements about the way a sung melody is transcribed. Such expert agreements may be taken as intersubjective valid results. But in our view, it is the shared cultural background, *i.e.*, the musical mind that coins conceptual hearing and comprehension. Thus, we see the primary advantage of computer tools in their function as an *external control* on the basis of acoustical criteria to conceptualise singing. Apart from this, this tool permits researchers to achieve reliable and valid data without interference from the musical mind, since analysis no longer relies on hearing only. For instance, by analysing a sung tune on the basis of each single syllable and its acoustical data, the culturally defined relations among the sounds, that usually affect our listening, can be eliminated. Although analysed as single events, as a consecutive series they yield a complex structure showing syllables with their pitches and their timing.

A Computer-Aided Method for Analysing and Representing Singing

In the following, we describe the new method we devised for analysing and representing pre- or unconventional singing. Although we exemplify this method within the context of an investigation into a young child's song-singing, it is also applicable to various other research problems that require musical analyses with respect to the organisation of pitch and time in a more differentiated, reliable, and valid way.

Any computer-aided method assumes (i) that the singing is recorded on an audio or video tape and (ii) that the computer is able to record and digitise audio data from a tape. In order to aid the analysing process, an appropriated software is necessary. It should be able to provide information on the timing and on pitch for certain singing events selected by the researcher. The first task can be done by virtually any sound recorder and editor, whereas the second task needs special software. Our solution is a combination of a commercially available sound recorder with a self-written program for pitch analysis. For more details see Appendix 0.1.

Figure 1 is a typical output of a pitch analysing program. The upper part shows the envelope of the sound. It gives some information on loudness. The lower part gives the raw pitch data. Only sounds that reach a certain loudness-threshold can be analysed.

This raw data has to be further processed for two reasons. First, the pitch analysing program offers too much data by showing all the wiggly details of the pitch curves. For

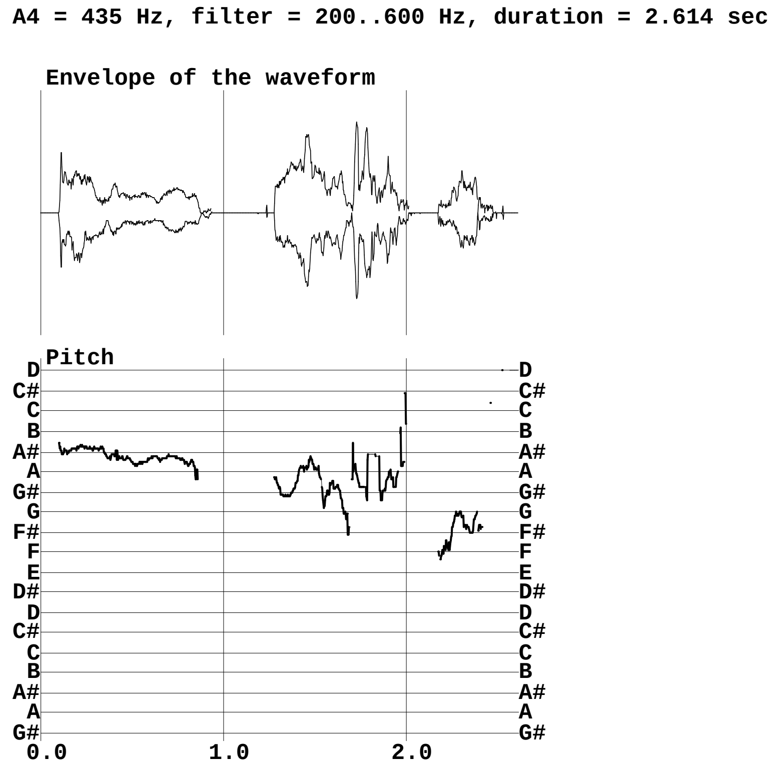


Figure 1: Example of a print-out of our pitch analysing program.

most research questions, the amount of information provided is too rich. Hence, the researcher has to choose a useful reduction of the information to a limited set of categories of pitch curves that is relevant to the research question. Second, the print-out of our program shows that certain important information is missing that would be necessary for a valid description of the data. Such information concern *e.g.* joint singing, breathing (to determine the phrases), the syllables of the lyrics, and the identification of the singer's voice.

In the context of our research on young children's song-learning processes, we come up with a list of categories represented by the symbols shown in table 1. The first five symbols define our categories with respect to the relevant parts of the pitch curves. The symbol `W` stands for syllables that are spoken and not sung. The symbol `X` means that it is not possible to gain reliable data on pitch from the computer tools because of a strong disturbance. The symbol `H` stands for 'help' and indicates that the syllable is sung by the researcher or presenter of the song, thus not by the child. Joint singing events are symbolised by an additional circle around the center of the symbol. The encircled symbol represents the louder of the two singer's production. Note that in other research contexts, *e.g.* an ethnomusicological one, other symbols and categories may be more appropriate.

An Example

As an example of this new method's outcome, Figure 2 shows our graphic representation of a girl's song-singing at age 4;5 years. It is an excerpt out of an acquisition process that occurred in a natural context of song learning. (For more detailed results obtained by this method, see *e.g.* Stadler Elmer, 1998, 2000a; 2000b; 2000c; Stadler Elmer and Hammer, in press). As the figure's title indicates, it is the girl's second solo performance of this song, and including all its previous events, this song counts as the eighth event. The learning situation is reflected be the fact that this event is initially guided by a trained female singer who instructed this song. We selected this example for illustrating such natural occurrences as is a transition from guided singing to joint singing, and finally to solo singing. Note that all this happened within a short period of time, namely 8.9 secs.

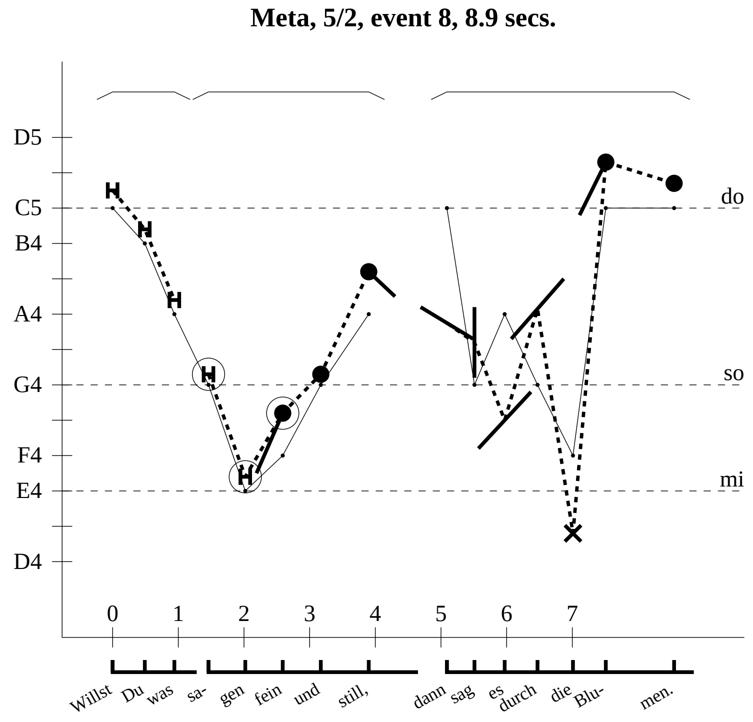


Figure 2: An example of the graphical representation of a song actually produced by a girl at age of 4;5 years. The title informs about (i) the singer, (ii) the song, (iii) the numbers of at least partially solo singings of this song by this singer, (iv) the number of times this song occurred while the child was present, and (v) the total duration in seconds. The legend of symbols is listed in Table 1. For more details, see the main text. Note that syllables 8-11 are the transcriptions of what is given (or not) in the raw data shown in Fig. 1.

Each syllable produced of the song's lyrics is categorised by one of the above-mentioned symbols. The horizontal and vertical position of a symbol's center define the time onset and the pitch, respectively. The vertical extension of the symbols 2-5 represents the pitch range of glissandi. Syllables belonging to the same phrase are connected by a dotted line. The end of a phrase is determined by breathing.

It should be emphasised that the symbols are placed in a time and pitch *continuum.* Therefore, they are not restricted to western categories even though the vertical axis has tics for the well-tempered scale (A4 calibrated at 435 Hz), and the horizontal axis counts the regular beats of the tics (here 120 beats per minute). These scales of the axes provide information on the western conventions. Where possible, the tonic triade (do, mi, so) is marked by dotted horizontal lines for a further orientation. Also, for orientation, the actual song model is drawn in thin lines connecting small dots. It stands for an idealised depiction of the pitch organisation that was repeatedly presented to the child. In order to reduce confusion, the timing of the song model's syllables is set identical with the timing of the syllables actually produced. Since the pitch organisation of the song model is given according to the well-tempered scale and serves as a normative account, one can easily see the pitch deviations of the syllables actually produced. For example, Figure 2 shows, that the strong deviation from the song model at the beginning of the third phrase corresponds to many (and sometimes large!) glissandi. Such large glissandi that lack a stable state with respect to pitch, are also observed by Sergeant (1994) in children's pitch matching with untrained singing voices.

Below the time axis, the onset of each syllable and ending of each phrase are shown in thick vertical and horizontal lines, respectively. Additionally, a transcription of the lyrics produced is shown. Here, one can easily see rubati or metric irregularities. In the present example, all syllables of the song model are quarter notes, except for "still", "Blu-", and "men" that count half notes.

Because the manual drawing of such graphical representations would be tedious work, we condense the whole information of Figure 2 into a small data file from which a special computer program produces the graphical representation. As an example, Appendix .1 shows the data file from which Figure 2 is made.

The detailed description shows that this girl is already familiar with the song model: She took over the entire song's lyrics correctly, and additional similarities are given in the song's first part where she had been given scaffold by the more competent singer. After the transition from joint to solo singing, she shows some adjustment to the model's melodic contour. In the second part, the similarity is obvious in the contour of the last three notes where she adapts to the model by exaggerating the interval of a fifth with a large leap upwards and by ending with the two closely pitched notes. Interestingly, the gross deviations from the model's melody occur at the beginning of the second phrase, and there we see even four symbols that indicate unstably produced pitch qualities. To gain a solid interpretation about the more psychological aspects of her song-learning process, we would need information at least about her previous and her subsequent singing of this song.

Utility of the New Method

In order to illustrate what the traditional method might yield with the same material represented above, we asked a professional musician to listen to the song-singing and to transcribe it. The result is shown in Figure 3. We deliberately did not give special instructions because our aim was to see the well-known effects of categorical perception with our material. It would be worth to study how special instructions and special training can improve analytical hearing of musicians.

Now let us compare the outcome of the traditional method with our method.

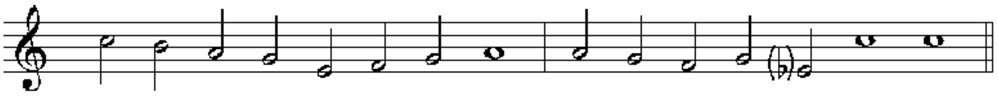


Figure 3: Musical transcription of the same vocal production on which Fig. 2 is based. The transcription was made by a professional musician who has been conducting children's choirs for many years.

The first five notes, where the presenter sung or dominated joint singing (see symbols H and joint singing), are fairly identical. In the following, taking into account the glissandi in Figure 2, the sixth and eighth note can be interpreted as roughly being within the same pitch category. Thus, we may say, that pitch representation of the first part of the song does not considerably differ between the two methods. Note, that the musician's perceived melody well suit to the song model, showing his understanding of the musical meaning. But the second part, where the child's sung pitches are less stable, we see considerable discrepancies between the computer-aided method and the musician's transcription. The first three notes illustrate the aforementioned musician's categorical perception, namely by adjusting the half-tone deviations to a key assumed to be F major. The fourth note even deviates approximately two semitones from the symbol gained by the aid of the computer analysis. Therefore, the musician's transcription proves to be wrong, even on the assumption that the key would be F major. The estimation of the disturbed signal (see symbol X) seems almost agreed, whereas the last two notes again differ clearly between the two methods.

With respect to the song's produced *temporal pattern*, the musician's representation shows that he perceived temporal categories of two values, namely, either one or two beats. In contrast, the representation gained by computer-aided analysis shows some temporal variations that would not be well represented by a two-value timing category system.

This brief comparison shows that the musician's transcription of pitch agrees with the outcomes of the new method as long as the sung melody suits the conventions, as is the case with the trained singing voice in the initial part of the song.

Since the child is producing this song only her second time (cf. perf. 2), we cannot expect her to perform the same quality as a competent singer. Note that the unstable part of her singing concerns just the one with the largest deviations in pitch from the song model. Yet, the musician's representation of her singing suggests an overall rather nice melody, except for one note where he was not sure. This reflects the well-known “categorical perception” (see introduction). When comparing to the newly proposed method, it is obvious that the traditional one does not account for the singer's still unstably produced pitches. Rather, the musician's use of the conventional music notation *a priori* limits the scope of perceiving and describing the phenomena.

Not only does the new method use more differentiated tools by locating symbols for pitch qualities in a continuous time and pitch co-ordinate system, but also these results are gained by combining analytic hearing with acoustic data given by the computer program. Thus, the basis for analysing and describing the phenomena is more reliable than pure auditory perception. Moreover, it allows to represent the organisation of three pertinent components (pitch, lyrics, time) of a sung performance as a structural entity, rather than as isolated parameters.

Concluding Remarks

Musical practice is communication, and written symbols play an important role as a medium for the conservation and transmission of musical meaning. The music notation system serves the composers to communicate musical meaning, and the notation is intended as guidelines for interpreting by musicians. Different from the artistic contexts and musical practice, scientific communication about actually produced melodies has to be as precise and unequivocal as possible. Akin to the discrepancy between written and spoken language, the sung melody requires a description that accounts for possible deviations from the idealised conventions represented by musical notation. These divergent criteria for coding and decoding musical meaning explains why we cannot expect the musical notation to describe adequately actual performances. What is lacking is some kind of symbolic system similar to the ones for speech to be represented by letters for reading, and by phonetic letters for representing spoken sounds.

As discussed in this paper, the main research problem with unconventional or preconventional singing is eliminating the culturally coined musical mind that tends to overestimate cultural concepts in perceived phenomena. Moreover, for certain research questions we need to go beyond evaluating single acoustic parameters provided by computers, but analyse and describe the organisation of larger behavioral units. In order to solve these problems, we devised the proposed method that uses acoustic measures for representing detailed structural aspects of song-singing that are more differentiated and reliable than music notation.

By proposing to combine auditory with acoustic analyses for the analysis of singing, we follow the approach already introduced by H. & M. Papoušek (*e.g.* 1981). Their microanalyses of musical elements in infants' and parents' vocalisations have not yet been seriously transferred to research in singing. A combined method uses a computer program that provides various acoustic measures on the sounds' pitches and timing. Together with listening, we obtain a computer-aided analysis that has clear advantages over the pure auditory analysis. For describing actually produced melodies, it allows to control and even eliminate the listener's musical mind while analysing a sung melody.

In addition to what Papoušeks' proposed, we emphasise using a more differentiated symbolic system for reducing the detailed acoustic information than by the traditional music notation system. The computer-aided detailed information make it possible to go beyond the given western musical categories of pitch and time by exploiting the entire continua of these parameters. By this, the culturally given category boundaries are dissolved. Nevertheless, a reduction of the detailed information obtained by the computer is still necessary, if we want to describe and understand more than a single and isolated parameter. However, any methods for reducing the acoustic data should be subordinate to the research questions. In our research context of children's singing development, we distinguish between several kinds of stable and unstable pitches and depict these categories with symbols that are placed in the co-ordinate system.

We illustrated our method by an example of a young girl's song-singing, and we demonstrated its utility by comparing its outcome with the one achieved by the musician's traditional method. This demonstrated shows the expected bias towards categorical perception, and moreover, the traditional method cannot account for unstably sung pitches nor for pitches in-between the given categories.

The example confirmed that the proposed method not only provides external control of perceived stimuli, but also allows to communicate the analysed data by a symbol system that is more adequate for describing singing than the conventional musical system and selected acoustic parameters.

In this way, it is now possible to study a series of interesting issues related to vocal musical expression and underlying cognition. First, short- or long-term changes in the structure of vocal musical expression as they adapt towards our western musical conventions, can be investigated. Second, in research contexts that focus singing in the light of learning and development, underlying organisational rules of vocal musical expression can be studied with respect to pitch and time. (For more details see Stadler Elmer, 1998; 2000a; 2000b; 2000c; Stadler Elmer & Hammer, in press). Furthermore, our method might be useful in ethnomusicological research.

Despite of these advantages, we are far from claiming that the new method would provide an ultimately true or objective description of what we observe or perceive. Although the paradoxical relation between knowledge and understanding on the one hand and perception or observation on the other, is reduced at one stage in the research process, it reappears elsewhere. The reason for that lies in the fact that we have to give meaning to the data computers provide. Hence, we need socio-cultural knowledge on the phoneomena. Interpretation already starts with distinguishing between the relevant and irrelevant data, e.g noise from sung sounds. More critical are problems to determine intended events from observed events in the interpretation of the data. How can we decide to what extent structures of a child's sung performances represent her or his mental representation or are destorted by a lack of vocal control? Such questions need to be addressed in subsequent research. Eventually, as with any microgenetic method (Catán, 1986; Siegler & Crowley, 1991), the proposed method is time-consuming. The researcher has to decide whether the research questions require gaining data at microanalytic levels or not.

Nevertheless, altogether this method provides a more reliable and differentiated access to vocal productions that for a naïve listener would appear "wrong" or "out-of-tune". It facilitates consideration of certain questions that could not be otherwise addressed, *e.g.* regarding structural changes across time in the song production of individuals, or investigations of relationships among a number of aspects of song production, learning, and development. *What* is sung has its own regularities to be discovered in future research.

Acknowledgments

This research method was devised within a project supported by the Swiss National Science Foundation (Nr. 11-37693.93 and Nr. 11-042221.94). The writing was made possible by a grant from the University of Zürich to the first author. We thank an anonymously kept musician for his transcription.

Description of Our Computer Tools

Here we describe in some details our computer tools and how we get timing and pitch information from our audio data. The software including a tutorial are available from the authors on request. Note that our tools, especially the self-written pitch analysing program, are optimised for analysing singing of young children.

For our study we use two different PC's (an old one with a 386 processor plus an arithmetic processor and Windows 3.1 as the operation system as well as a newer one with a Pentium processor and Windows 95). Both PC's are equipped with soundcards. Because the sound recorder delivered with the operation system is rather limited in its functionality we use a commercial (shareware) sound recorder and editor (CoolEdit V1.31) which is widespread. Our self-written pitch analysing program communicates with the sound recorder by constantly listening for data sent by the sound recorder to the clipboard (*i.e.*, a part of the memory accessible for all running applications).

Description of the Sound Recorder

In order to extract information on time and pitch, the following features of CoolEdit are important:

- On the screen the sound signal is drawn as a function of time. From this one gets immediately the timing of the onset of the syllables that are sung staccato because the syllables are clearly separated by silence.
- Parts of the recorded data can be played back between freely selected time points. This is important for obtaining information on the timing of syllables which are not clearly separated by silence.
- A selected part can be copied onto the clipboard. This is the way to deliver data to our own pitch analysing program.
- A large variety of manipulative functions (amplification, designable filters, etc.) is available to improve the signal-to-noise ratio.
- A spectral analysis of the data can be obtained either as a curve of spectral intensity *versus* frequency for selected time points or as a color-coded sonogram of the whole data. We use the spectral analysis as a complementary method in cases where our self-written pitch analysing program does not yield reliable data.

It should be remarked that these features may be also available in other commercially available sound editors and recorders.

Description of the Self-Written Pitch Analysing Program

Extracting the pitch from an audio signal means to find the fundamental frequency (F_0) of a periodic signal. In the literature, a huge variety of methods has been reported for this problem. For an overview of the most important ones, see the monography by Hess (1983). Basically, the methods can be divided into two subgroups, spectral methods and temporal methods.

A *spectral method* is based on the calculation of the *power spectrum.* Any audio signal can be thought of as the sum of sinusoidal oscillations of different frequencies. The power spectrum gives the strength of oscillation for each frequency. The spectrum of a strictly periodic signal (which is not necessarily sinusoidal) contains only the oscillation with the fundamental frequency and/or oscillations with integer multiples of the fundamental frequency (*i.e.*, the harmonics). Graphically such a spectrum is characterised by very sharp peaks appearing equidistantly. Note, that the fundamental frequency is often not present (*e.g.* in the voice of an opera singer with its typical formant around 3000 Hz.). But the distance between two consecutive higher harmonics also defines the fundamental frequency and therefore pitch. Nonperiodic and noisy signals are characterised by relatively flat and broad spectrum. A power spectrum is usually calculated for a finite interval at an arbitrary point on the time axis. A sonogram is a graphical presentation of the spectra of all time points. The x-axis is the time axis whereas the y-axis is the frequency axis. The intensity of an oscillation of a certain frequency at a certain time is coded by a grey scale or a color code. Periodic signals manifest itself typically by the appearance of stripes parallel to the time axis. The averaged distance between the stripes is a measure of the pitch. A vibrato, *e.g.*, leads to wiggly stripes.

The sound recorder we use is able to perform spectral analysis and shows sonograms. But it is a very tedious task to obtain quantitatively the pitch from this spectral analysis as a function of time. Therefore, we use its capacity to perform spectral analysis only for a comparison with uncertain results or in cases where our pitch analysing program fails to yield reliable results.

Temporal methods do not perform any spectral analysis. The main reason for avoiding a spectral analysis is to save computation time. Temporal methods can be very fast because they usually need only integer arithmetics. For this reason we have chosen a purely temporal method that should run fast even on an older PC without a numeric processor. An algorithm, working within the time domain, has to look for some characteristic features which reappear periodically. We have chosen the absolute minimum of the waveform or alternatively the absolute maximum.

The fundamental frequency is given by the inverse of the time difference between two consecutive events of the characteristic feature. The main problem with a temporal method is to overcome the confusion between characteristic features that seem to be similar. This is especially important for periodic signals where the fundamental frequency does not appear. Such a signal is characterised by a waveform that resembles a landscape with many hills and valleys having different heights and depths. But eventually, the valleys and hills of the same depths and heights reappear defining the fundamental frequency. One of the strategies we used in our algorithm to tackle this problem is a restriction of the expected pitch to a certain interval given by the user of the program.

Since we are analysing the singing of young children, this problem is not a very serious one because in these signals the fundamental frequency is usually present. Nevertheless, with this kind of data the program sometimes computes the wrong octave.

It should be noted, that all pitch analysing methods have serious problems of analysing polyphonic signals. Thus, getting the pitch of two singers singing with roughly the same loudness is almost impossible.

After the analysis, the waveform of the data as well as its pitch pattern are shown graphically on the screen as a function of time. The program calculates and shows pitch only when the amplitude of the signal is larger than a certain threshold. If the program does not succeed in calculating the pitch, it draws the last successfully calculated pitch in red. This may lead to red horizontal lines which can be taken only as a very crude approximation of the actual pitch (if it exists). The pitch curve is shown together with horizontal lines that mark tempered pitches based on a freely definable calibration of the middle A4 that can be set in the range between 430 to 450 Hz. By moving the mouse cursor, one gets the pitch in Hz as well as the nearest note on the tempered scale and its deviation in cents for any point on the time scale. The result can be either printed or stored as a list of readable data on the hard disk.

Figure 1 gives an example of a print-out of our pitch analysing program. The heading text line shows (i) the frequency on which the tempered scale in the lower part is based, (ii) the expected pitch range (given in Hz), and (iii) the duration of the sample. The upper plot shows the envelope of the waveform. The horizontal line in the middle means no wave, *i.e.*, silence. The pitch is shown below the envelope. Note, that pitch is only calculated when the distance between the upper and lower envelope is large enough, *i.e.*, when the sound is loud enough. When the loudness occasionally drops below the threshold, especially just before the end of a syllable (here *e.g.* just before the 2-sec. line and near 2.5 sec.), the results may be spurious. The thin and short horizontal line near 1.82 sec. means that the program could not calculate a reliable value for the pitch. Therefore, as explained above, it draws the last successful calculation. With unclear or equivocal sounds, it is useful to apply spectral analysis as a complementary method.

Example of the Data Representation of the Singing Structure

The ASCII-text file from which Figure 2 is drawn reads as follows:

```
Mathy, song 5, perf. 2, event 8, 8.9 secs.
3
3
 8 24.5 24.5  0.04 Willst
 8 23.4 23.4  0.53 Du
 8 21.4 21.4  0.98 was
5  1.32
18 19.3 19.3  1.50 sa-
18 16.4 16.4  2.06 gen
13 16.5 18.2  2.63 fein
 1 19.3 19.3  3.21 und
 2 22.2 21.5  3.94 still,
7  4.69
 4 21.2 20.3  5.13 dann
 5 21.2 19.2  5.55 sag
 4 17.2 18.8  6.01 es
 4 20.3 22.0  6.51 durch
 7 14.8 14.8  7.05 die
 3 23.8 25.3  7.55 Blu-
 1 24.7 24.7  8.59 men.
   8.89
```

The first line is the title. The number in the second line counts the number of phrases (here 3 phrases). Each phrase is headed with a line containing the number of syllables and the ending time (in secs.) of the previous phrase. The line of each syllable contains (in this sequence) the symbol code (see Table 1), two numbers carrying information on pitch, the onset time (in secs.), and the syllable of the lyrics. The very last line contains only the ending time of the last phrase.

Information of pitch is coded in the following way. The number before the decimal point is the pitch on the well-tempered scale where zero denotes C3. The digit behind the decimal point counts the (upper) deviation in cents divided by ten. Thus 17.2 means F4 plus 20 cents, whereas 15.7 means E4 minus 30 cents. The symbols 2-5 and 12-15 require two different numbers of pitch information because they denote the starting and ending pitch, whereas this distinction is not necessary for all the other symbols. Therefore, starting and ending pitch coincide and are represented by two times the same value (as an example, see the first 5 syllables).

References

Anderson, J. D. (1991). Children's song acquisition: An examination of current research and theories. The Quarterly Journal of Music Teaching and Learning, II(4), Winter, 42-49.

Atterbury, B. W. (1984). Children's singing voices: A review of selected research. Council for Research in Music Education, 80, 51-63.

Beilin, H. & Pufall, P. (eds.) (1992). Piaget's theory: prospects and possibilities. Hillsdale, N.Y.: Lawrence Erlbaum Ass.

Catán, L. (1986). The dynamic display of process: Historical development and contemporary uses of the microgenetic method. Human Development, 29, 252-263.

Davidson, L., McKernon, P.E., & Gardner, H. (1981). The acquisition of song: A developmental approach. Documentary Report of the Ann Arbor Symposium: Applications of psychology to the teaching and learning of music (pp. 301-317). Reston, Virg.: Music Educators National Conference.

Davidson, L. (1985). Tonal structures of children's early songs. Music Perception, 2(3), 361-374.

Davidson, L. (1994). Songsinging by young and old: A developmental approach to music. In R. Aiello (ed.) Musical Perceptions (pp. 99-130). New York: Oxford University Press.

Davies, C.V. (1986). Say it till a song comes. British Journal of Music Education, 3(3), 279-293.

Davies, C.V. (1992). Listen to my song: a study of songs invented by children 3-13. British Journal of Music Education, 3(3), 279-293.

Fernald, A. & Simon, T. (1977). Analyse von Grundfrequenz und Sprachsegmentlaenge bei der Kommunikation von Muettern mit Neugeborenen. Forschungsberichte: Institut f. Phonetik u. sprachliche Kommunikation der Universitaet Muenchen.

Fernald, A. & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. Developmental Psychology, 20(1), 104-113.

Fernald, A.& Kuhl, P.K. (1987). Acoustic determinants of infant preference for motherese speech. Infant Behavior and Development, 10, 279-293.

Fernald, A., Taechner, T., Dunn, J., Papoušek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. Journal of Child Language, 16, 477-501.

Flowers, P.J. & Dunne-Sousa, D. (1990). Pitch-pattern accuracy, tonality, and vocal range in preschool children's singing. Journal of Research in Music Education, 38 (2), 102-114.

Garnica, O.K. (1977). Some prosodic and paralinguistic features of speech to young children. In C.E. Snow & C.A. Ferguson (Eds.), Talking to children: Language input and acquisition (pp. 63-88). Cambridge: Cambridge University Press.

Hart, J.'t., Collier, R., & Cohen, A. (1990). A perceptual study of intonation: an experimental-phonetic approach to speech melody. Cambridge: Cambridge University Press.

Helfrich, H. (1985). Satzmelodie und Sprachwahrnehmung. Psycholinguistische Untersuchungen zur Grundfrequenz. Berlin: W. de Gruyter.

Hess, W. (1983). Pitch determination of speech signals: Algorithms and devices. New York: Springer.

Imberty, M. (1996). Linguistic and musical development in preschool and school-age children. In I. Deliege & J. Sloboda (eds.) (1996). Musical beginnings. Origins and development of musical competence. Oxford University Press.

Kelley, L. & Sutton-Smith, B. (1987). A study of infant musical productivity. In Peery, J.C., Weiss Peery, I., & Draper, T.W. (eds.), Music and child development (pp. 35-53). New York: Springer.

McKernon, P.E. (1979). The development of first songs in young children. In D. Wolf (ed.) New directions for child development, Vol. 3 (pp. 43-58). San Francisco: Jossey-Bass Inc.

Moog, H. (1967). Beginn und erste Entwicklung des Musikerlebens im Kindesalter. Ratingen: Henn (Orig. 1963).

Moog, H. (1968). Das Musikerleben des vorschulpflichtigen Kindes. Mainz: Schott.

Papoušek, M. (1981). Die Bedeutung musikalischer Elemente in der frühen Kommunikation zwischen Eltern und Kind. Sozialpädiatrie in Praxis und Klinik, 3 (9), 412-415, 3 (19) 468-473.

Papoušek, M. (1994). Vom ersten Schrei zum ersten Wort. Vorsprachliche Kommunikation zwischen Mutter und Kind als Schrittmacher der Sprachentwicklung. Bern: Huber.

Papoušek, M. (1995). Origins of reciprocity and mutuality in prelinguistic parent-infant 'dialogues'. In I. Markova, C. Graumann, & K. Foppa (eds.) Mutualities in dialogue (pp. 58-81). Cambridge: Cambridge University Press.

Papoušek, M. (1996). Intuitive parenting: a hidden source of musical stimulation in infancy. In Deliège, I. & Sloboda, J. (Eds.) Musical beginnings. Origins and development of musical competence (pp. 88 - 112). Oxford: Oxford University Press.

Papoussšek, M. & Papoussšek, H. (1981). Musical elements in the infant's vocalisation: Their significance for communication, cognition, and creativity. Advances in Infancy Research, 1, 163-224.

Papoušek, M. & Papoušek, H. (1989). Forms and functions of vocal matching in preconpremonical mother-infant interactions. First Language, 9, 137-158.

Seashore, C.E. (1938). Psychology of music. New York: McGraw-Hill.

Sergeant, D. (1994). Towards a specification for poor pitch singing. In G. Welch & T. Murao (eds.) Onchi and singing development (p. 63 - 73). London: David Fulton Publ.

Siegel, J.A. & Siegel, W. (1977a). Absolute identification of notes and intervals by musicians. Perception & Psychophysics, 21 (2), 143-152.

Siegel, J.A. & Siegel, W. (1977b). Categorical perception of tonal intervals: Musicians can't tell sharp from flat. Perception & Psychophysics, 21 (5), 399-407.

Siegler, R.S. & Crowley, K. (1991). The microgenetic method. A direct means for studying cognitive development. American Psychologist, 46 (6), 606-620.

Smith, B.L. & Kenny, M.K. (1998). An assessment of several acoustic parameters in children's speech production development: longitudinal data. Journal of Phonetics, 26, 95-108.

Smith, L. (ed.) (1996). Critical readings on Piaget. London: Routledge.

Stadler Elmer, S. (1990) Vocal pitch matching ability in children between four and nine years of age. European Journal for High Ability, Vol. 1, 33-41.

Stadler Elmer, S. (1996). Die Entwicklung des Singens: Eine kritische Diskussion der Beschreibungs- und Erklärungsansätze. Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 28 (3), 189-209.

Stadler Elmer, S. (1997). Die Anfänge des musikalischen Erlebens und Erkennens. In J. Scheidegger & H. Eiholzer (Hg.). Persönlichkeitsentfaltung durch Musikerziehung (S. 35-49). Aarau: Musikedition Nepomuk.

Stadler Elmer, S. (1998). A Piagetian perspective on singing development. Jahrbuch der Deutschen Gesellschaft für Musikpsychologie, Bd. 13, 108 - 125. Göttingen: Hogrefe.

Stadler Elmer, S. (2000). Liedersingen mit Kindern: Strukturgenese im sprach-musikalischen Ausdruck. In S. Hoppe-Graff & A. Rümmele, A. (eds.) Entwicklung als Strukturgenese.

Stadler Elmer, S. & Hammer, S. (1999). Sprach-melodische Erfindungen einer 9-jährigen. Research Report, University of Zuerich, submitted for publication.

Stern, W. (1914/65). Psychologie der frühen Kindheit. Heidelberg: Quelle & Meyer.

Sundberg, J. (1982). Perception of singing. In D. Deutsch (ed.) The psychology of music (p. 59-98). New York: Academic Press.

Tohkura, Y., Vatikiotis-Bateson, E., & Sagisaka, Y. (eds.) (1992). Speech perception, production and linguistic structure. Washington : IOS Press.

Trehub, S.E., Unyk, A.M., Kamenetsky, S.B., Hill, D.S., Trainor, L.J., Henderson, J.L., & Saraza, M. (1997). Mothers' and fathers' singing to infants. Developmental Psychology, 33 (3), 500-507.

Welch, G.F. (1994). The assessment of singing. Psychology of Music, 22, 3-19.

Werner, H. (1917). Die melodische Erfindung im frühen Kindesalter. Bericht der Kaiserlichen Akademie, Wien, 182, 1-100.

Addresses of correspondence:

PD Dr. Stefanie Stadler Elmer
Dept. of Education
University of Zürich
Rämistr. 74
CH-8001 Zürich
e-mail: stadler@paed.unizh.ch

PD Dr. Franz-Josef Elmer
Institute of Physics
University of Basel
Klingelbergstr. 82
CH-4056 Basel
e-mail: Franz-Josef.Elmer@unibas.ch

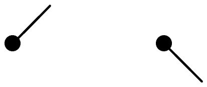
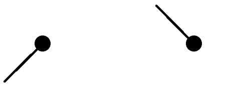
Code	**Symbol**	**Description**
1	●	Stable pitch
2		Stable pitch, ending with upward or downward glissando
3		Stable pitch, starting with upward or downward glissando
4	╱ ╲	Unstable pitch, but clear upward or downward glissando
5	|	Unstable pitch with glissandi in any direction and/or unidentifiable, fuzzy pitches within context of singing (prolonged vowel)
6	W	Pitch of a spoken syllable
7	╳	Estimation on the basis of disturbed signales
8	H	Syllable sung by the researcher
+10	○	Joint singing

Table 1: Table of symbols. In the case of joint singing, a circle is drawn around the center of the symbol that is the geometrical center except for symbol 2 and 3 where it is the dot. In the graphical representation of a song actually sung, the x and y coordinates of the position of the symbol's center denote the onset time and the pitch, resp., of the corresponding syllable.