Visualizing Music Structure through Animation: The Making of the Film *Head Voice, Chest Voice*

Hugo Zemp

While many ethnographic filmmakers, as well as ethnomusicologists interested in audio-visual means, are concerned about refining methods for the decoding and interpretation of visual texts, this article proposes a different perspective: that of visually reproducing in a film on musical features the results of research obtained with other visual means. An unexpected outcome was the discovery, thanks to the filmmaking process, of an important musical feature which has not been noticed and analyzed so far.

Animation in Ethnographic and Ethnomusicological Films

Despite the possibilities animation has to express through visual means complex structures or processes which are difficult or impossible to show with footage shot on location or to explain with narration alone, it has not yet been widely used in ethnographic and ethnomusicological films. This contrasts with its frequency both in educational films and in TV broadcasts on "hard" and medical sciences. Without doubt, one reason is its high cost (educational films and TV broadcasts on hard sciences usually have larger budgets than ethnographic films!). Another reason might be the reluctance of many anthropologists and ethnomusicologists to invest much time and energy in fund raising and assimilating the complex technology of animation.

I know of two ethnomusicological films using simple animation of diagrams to visualize and explain construction of musical instruments. In the film *Le Mouet* [1972], African filmmaker Moïse Zé intercuts actual footage and animated diagrams to show the making and tuning of a harpluth in Cameroon. In my film, *Shaping Bamboo* [1979], about panpipes of a Solomon Island society, two animation sequences make it possible to visualize the order in which instruments of different sizes are made, and the principle of their octave tuning, thus avoiding repetitive showing of the same technical processes used for each instrument of the orchestra. To my knowledge,

_Hugo Zemp_, ethnomusicologist, the Centre National de la Recherche Scientifique and affiliated with the Département d’ethnomusicologie, Musée de l’Homme, Place du Trocadéro, 75116 Paris, France. Zemp carried out field work in West Africa (Ivory Coast), Oceania (Solomon Islands), and Switzerland. He made several films following Jean Rouch's precept that the researcher should act as cameraperson, working with a native soundperson.
only one ethnomusicological film has so far tried the more ambitious task of explaining elements of musical structure through visual means: *Mbira: The Technique of the Mbira dza Vadzimu* [1976], by filmmaker Guy Zantzinger and ethnomusicologist Andrew Tracey. It uses the freeze-frame technique and superimposed animation of letters over actual shots to demonstrate rhythmic and harmonic patterns of playing a South African lamellophone. Its harmonic analyses are especially complex, and in order to follow the demonstration, the spectator needs solid knowledge of methods used in Western musical analysis. In my latest film *Head Voice, Chest Voice*, my aim was to visualize characteristic musical features in such a way that an audience without extensive training in Western classical music theory could also follow and understand the principles of structure and performing of a kind of music characterized by a specific voice technique: the yodelling of an Alpine region.

**The Series “Jüüzli” of the Muotatal**

In watching and analyzing films on traditional music, including my own previous films on Solomon Island music, I often become frustrated by either a narrow focus on limited aspects of music making, or on the contrary, by a superficial treatment of a wide range of situations and facts. It is simply not possible in a single film to provide a holistic approach to several musical genres of a people while simultaneously presenting extensive music performance without overloading the film with both visual and aural information contained in the actual footage, not to mention the risk of overloading it with heavy narration. If the emphasis is put on lengthy filming of musical performances, one may regret the thinness of filmed ethnographic contextualization, and *vice versa*, while musical analyses are practically always absent.

To attempt to solve the dilemma, I designed a series of four films, each presenting the same musical genre from a different, complementary perspective. Detailed discussions of shooting and editing strategies on the one hand, of ethical issues on the other, are given elsewhere [Zemp 1988a, 1988b]. Each film can be viewed independently, but for a better understanding and global vision of this music, which has major functions for local and national identity, it is recommended to see the films in the following order:

The first film, *Yootzing and Yodelling* [1986], deals with the situation of the local musical tradition confronted with institutionalized national folklore. Besides presenting performances in the various social contexts, the film expresses the native view through filmed conversations.

The second film of the series, *Head Voice, Chest Voice* [1987], which is discussed in this paper, presents in contrast musical analyses from the ethnomusicological point of view, explained through spoken commentary.

The third film, *The Wedding of Susanna and Josef* [1986], and the fourth,
Glattalp [1986], respectively concern a wedding, and going up to the Alpine
pasture. Spoken explanations are unnecessary since these single-day events
are easy for the spectator to follow and the native views expressed in the first
film, as well as the musical analyses in the second one can be referred to.

This film series is the result of field work I carried out since 1978 in the
Muotatal, a small valley of central Switzerland, and by my native soundman
and collaborator, also a performer, who studied the local tradition for his
teacher's diploma. Shooting on location took place during several periods in
1983 and 1984. The second film of the series, because it required further
consideration of structure and animation, was completed last, in November
1987.

The Characteristic Musical Features

Previous research on Alpine yodelling and of the local tradition of the
Muotatal [Baumann 1976; Betschart 1981; Sichardt 1939; Zemp 1979], has
shown the following main musical features:

a) register change between head voice and chest voice;

b) use of meaningless syllables, containing mainly the vowel “U” in head
voice, and the vowel “O” in chest voice;

c) singing with tense voice by traditional performers, characteristically
revealed at the ends of the melodic lines where the last tone is cut with a
glottal stop or with a rapid falling glissando, while yodlers with schooled,
relaxed voices hold the last tone on an even pitch until it dies out;

d) raising of the fourth degree of the diatonic scale, called “Alphorn-fa”
since it corresponds to the eleventh harmonic (a raised F if the fundamental
tone is C) of the natural scale used by the Alphorn, and which is considered
an archaic feature of the repertoire;

e) neutral intonation of the third and seventh degrees of the diatonic scale
by traditional performers, tones which are considered off-key by followers
of schooled yodelling, as well as by music teachers;

f) pieces with a traditional form of two different melodic lines, each
repeated, giving rise to the structure AABB;

g) polyphonic organization of two-part and three-part yodelling, where
the second voice-part always moves in the higher chest register, sometimes
crossing the leading voice, while the third voice-part remains in the bass
range, mainly on two tones corresponding to tonic and dominant in har-
monic analysis.

All these features can be described and analyzed in a written text supple-
mented with musical examples printed in Western staff notation. If the
research were exclusively for an audience of ethnomusicologists, then mak-
ing the film Head Voice, Chest Voice, would not have been necessary. In fact,
the printed notes accompanying the record which I published [Zemp 1979]
already contain most of these analyses (although without musical notations), which are thus easily available to ethnomusicologists. But experience has shown that natives of the Muotatal, though they know the record and widely discuss it, did not read the text,² no doubt because the long text in small print does not look attractive to them, because its content seems too technical, and/or simply because they are used to putting a record on the record player to enjoy the music and not to read lengthy record notes as if they were a school book.

To reach this public from the Muotatal as well as a wider audience from elsewhere not familiar with reading musical analyses, I searched for a way to visualize these features and chose as the best solution linear graph notation, used by some ethnomusicologists for many years in printed material, but which I had the idea to animate synchronously with sound.

Overall Structure of the Film

One of the strongest critics of didactic educational films, Colin Young [1976:68], wrote that the filmmaker “kept all the aces, controlling the flow of information and letting us see only what he wanted and what fitted his story or his thesis.” In short: “a take-it-or-leave-it option.” Of course, but is this always necessarily bad?

The researcher, even if he is also a filmmaker, has to assume responsibility for what he does. Does he want his films to be useful to the people with whom he has worked, as well as elsewhere, to promote a wider understanding of and respect for other musical traditions, or are his films intended mainly for the narrow circle of visual anthropologists, and inside this circle for those who defend exclusive observational cinema? Almost everywhere in the world official music education in school is based on Western academic music training. School teachers know nothing about and/or deprecate local music traditions. Ministries of Education rarely publish school material on traditional music. For ethnomusicologists, making the result of their research useful to the people whose music they study is an ethical necessity [Zemp 1988b]. I do not know of any way more effective than a didactic film allowing one to hear and to see, and thus to recognize and understand musical features.³ The problem is not to condemn didactic films in favoring observational cinema, but to avoid the justified reputation of boredom which many didactic films have, to make them interesting and, why not, entertaining.

In this series about Swiss yodelling, cultural contextualization is largely provided in three films, and thus it was possible in the one we are discussing here to concentrate on purely musical aspects. To maintain the interest of the spectator, I tried to vary the size, color, and layout of the animation on the screen, although this variety is never gratuitous, but helps to strengthen the
perception of the musical features. Also, since the sequences of animation should not be too long, a balance between shots of actual performance, animation, and other shots had to be found. Furthermore, I wished to avoid dehumanizing the performers and using them as anonymous beings to demonstrate abstract musical features. Therefore, the audience is reminded with one or two short shots of the situations in which most of the performers had been seen in the first film.

Short spoken commentaries recall their names and explain the main musical features visualized in the graphic animation. In all my films I systematically avoid voice-over narration during a musical performance; in the animation, the commentary is heard on freeze frames between musical fragments, or synchronously on blinking markers which visually support the spoken explanations. Also, since experience shows that voice-over narration either distracts from the images, or that paying attention to images makes it difficult to follow sophisticated narration, I chose images in the sequences filmed on location where "nothing was happening" to place the commentary. The editor would not have kept some of these shots or would have shortened them if I had not insisted on keeping enough footage to place the comments.

Real Time Graphic Linear Transcription

As an alternative to Western conventional music notation, linear handmade graphs have sometimes been used by ethnomusicologists. The advantage is that it is accessible to readers who do not read conventional staff notation.4

Why hand graphs and not automatic transcriptions? To my knowledge there exists no method for real time automatic linear transcription of a melodic line, and if there is one, we at our ethnomusicological research group in Paris did not have one at our disposal. What we did have was an old model of a sonagraph transcribing 2.4 seconds or 4.8 seconds of sound spectrum. It would have been possible to make a collage of many short fragments to obtain a sonagraph transcription of a whole piece, and then to animate it frame by frame synchronously with music. But the information on a sonagram—even if an enlargement enhances the lower spectrum and thus the fundamental pitches—is much too dense to be read in real time. In the case of three-part polyphony such as in some yodelling, the interpretation of a single sonagram of 4.8 seconds of music would probably require more than half an hour with an expert in spectral analysis. The movements of the different voice-parts have to be distinguished, the overtones have to be separated from fundamental pitches, and the reading is further complicated by traces of external noise (cow bells in our case) which are inevitably on a sound recording if it is not made in a studio.5

Conversely, hand graph animation makes it possible to simplify and to
select the information. The characteristic musical features mentioned before thus could successively be visualized, one at a time. To achieve the final animation, the following steps were necessary:

Step 1: The filmed yodels were transcribed by ear in conventional staff notation and analyzed. In some cases I had chosen to film particular pieces because of their specific features which I had analyzed several years before.

Step 2: The pitches were measured with a Stroboconn, an engine provided with twelve windows corresponding to the twelve notes of the chromatic scale. A main knob allows one to set the turning stroboscopic discs and thus to measure the pitch deviation from the closest tone of the tempered scale. From the figures obtained in cents (a cent is a hundredth of a semitone), the size of the intervals could be calculated, showing, for example, that in traditional yodelling, the neutral third degree is approximately 350 cents distant from the fundamental tone of the scale, that is an interval between a major third (400 cents) and a minor third (300 cents) of the tempered scale.

Step 3: The staff notations of entire yodels were transposed to linear hand graphs. Checking of the tape recordings made it possible to localize the register break in order to draw the horizontal separation line. To visualize in detail certain musical features, enlargements of shorter musical fragments were drawn.

Step 4: To facilitate counting of frames for each change of pitch, an optical sound track was made. On an editing table with two super-16mm film screens, it was possible to observe on one screen the lip movements of the singers, and on the other screen, the changing configurations of the optical sound track. Thus the pitch changes could be localized with a precision of one or two frames. Watching this sound track, I discovered strange and strong impulses lasting five to eight frames at the passages from head voice to chest voice in yodels of traditional performers. Something was happening to which I had never payed attention, since I had searched until then mainly for the exact size of intervals in neutral intonation. To analyze this phenomenon, sonagrams of these passages were made.

Step 5: The examination of the sonagrams showed that traditional performers reinforce the register break in quickly raising the pitch before falling from head voice to chest voice [Figure 1], while performers with schooled voices try to obtain smooth register transition [Figure 2]. Thanks to the use of an optical sound track for easier frame counting, I discovered a main musical feature which distinguishes traditional from schooled yodelling, and which the animation could visualize very clearly [Figures 3 and 4]. Besides this, sonagrams also made it possible to study the particular endings of melodic lines: soft dying out of the final tone by schooled yodlers [Figure 5], glottal stop [Figure 6], and falling glissando [Figure 7] by traditional performers. The time duration could be measured afterwards in millimeters and converted into frame numbers, while the ambitus of the
glissando could be calculated in terms of interval size, all the results being transferred to the hand graphs for the animation [Figure 8].

Step 6: The hand graphs provided with frame numbers were given to the animator who, with graphic software on a personal computer, redrew the lines on a grid. Blown-up printouts were made for easy checking, with the different voice-parts printed with different textures [Figure 9]. After proof-reading, separate Kodalith printouts (black background with transparent lines) for each voice-part were made.

Step 7: With the technique of superimposition, each Kodalith, lighted from below through a color filter, was filmed frame-by-frame, as a mask gradually revealed the lines from left to right. My first choice of color was red for solo yodels and first voice-parts of polyphonic yodels, yellow for the
second part, and green for the third part. The tests showed that yellow lines on the black background were brighter than red lines, so we inverted the colors since the second voice-part should not visually dominate the first part. But in polyphonic yodelling, where at times the first and the second voice-parts join in unison, the red line became “eaten” by the yellow and the visual impression wrongly suggested the disappearance of the second voice-part. After several tests we finally chose yellow for the first part, blue for the second, and red for the third.
Real Time Numerical Notation

To visualize the neutral intonation and the raised fourth degree (Alphornfa), I supplemented the graph in one blown-up fragment of a solo yodel with numerical notation [Figure 10]. Each numeral stands for a degree of the diatonic scale, the conventional symbol of an arrow indicating that the degree is lowered or raised approximately a quarter tone in regard to the major scale. In another yodel for which I wanted to show the neutral intonation for the entire piece without repeating a graphic representation, I superimposed numerals on a facial close-up shot of the singer. Here the numerals stand on an invisible horizontal line; dots over numerals indicate a move to a higher octave. According to the formal structure of the yodel (AABB), the numerals successively appear and stay until the end of each part [Zemp 1988a: Figure 25].
Figure 7. Sonagram showing the falling glissando of the last tone of a melodic line by a traditional performer (Alois Suter).

Just before this yodel in the film, the dance tune played on the diatonic accordion had the same melody. To show the similarity in form (although instrumental pieces are longer and thus have more repetitions of the different parts), superimposed letters appear synchronously and stay on the screen to show the overall formal structure of the piece [Figure 11].

These animated numerals and letters were filmed on black-and-white film to be superimposed, together with titles and subtitles, on each new print made by the film laboratory.

Humor

I believe that even in a didactic film (and especially in a didactic one), humor should not be absent. It helps the audience to relax, to be briefly entertained, and thus to be ready for another sequence which requires more concentrated observation. For example, after a one-minute facial close-up of a singer with superimposition of synchronous cipher notation which demands some attention from the spectator, I cut unexpectedly to a wider frame showing that during his whole performance, the peasant was sitting in front of his house door with a goat standing beside him [Zemp 1988a: Figures 25 and 26]. As I expected, the audience usually reacts with laughter. Another wink was less successful: after a shot showing the herdsman who watched and commented on his video image in the first film, I cut to his daughter stroking a rabbit and then the camera pans to the empty rabbit hutch—a transformed TV receiver. No spectator reacted during screenings or spoke to me about this shot; perhaps it is too short for the diversion of function to be easily recognized.

On the other hand, the audience reacts with laughter several times during the animation, where I did not expect it. At the first occurrence of the stylistic ornament which reinforces the register break from head voice to
chest voice [Figure 1]; at the first occurrence (and sometimes later) of the glottal stop at the end of a melodic line [Figure 8, middle]; at the first occurrence (and sometimes later) of the falling glissando at the end of a melodic line [Figure 8, bottom]; and in several cases of specific features of traditional performing technique which are not consistent with the aesthetics of the Swiss Yodler Union.

In the first two cases, the two musical features which have not been heard and seen so far are briefly explained over freeze frames; then a blinking star announces the ornamentation or the glottal stop several seconds before its appearance. When after a short suspense these two musical features appear, sound and images match so nicely that the audience enjoys it. In the case of the final glissando, the graphic line synchronously falling with the voice looks amusing, although according to my experience, nobody ever laughed at these places when only listening to sound recordings. In the cases where I wanted to express visually that specific traditional features are not allowed
in polished, folkloristic yodelling, I had the idea to cross out their graphic representation with large crosses: this visually expeditious method (consistent with dogmatic refusal) appears to have a hilarious effect.

When I first thought about the design of the animation, I was anxious not to bore the public with austere graphics. Although I did not plan it, I am happy that the audience reacts with laughter at several places. As I noticed from discussions with the public after the screenings, the passages which are perceived as funny sharpen the perception. This is certainly much stronger than dry explanations would be (my explanations are dry, but they become funny thanks to the graphic reinforcement), and the visual impact makes the spectators think about the serious issues which are being presented.

---

Figure 9. Blown-up computer printout of a hand graph showing the last notes of a yodel performed in three-part polyphony. For proofreading, the different voices are printed with different textures. The vertical grid indicating the pitches and the horizontal grid indicating the frame numbers are not visible in the final animation.

Figure 10. Computer printout of a hand graph supplemented with figures for the degrees of the scale. The arrows indicate that the tones are approximatively a quarter tone higher or lower than the closest tone of the tempered major scale.
Conclusion

According to their comments after viewing the film, spectators not familiar with musical analysis appreciated that for the first time they could see how music was structured, and how it developed over time. Even ethnomusicologists (although the film was not primarily made for them) told me that thanks to the animation they would now listen differently to this kind of music. I must add that for myself, who had to listen so many times to the sound recordings in order to make the musical transcriptions and the analyses, the fact of seeing the finished animation enhanced my understanding of musical structure and performance technique.

In his book, Ethnographic Film, Karl Heider concludes a paragraph entitled “Ethnographic Film Must Exploit the Visual Potential of Film” with the laconic sentence: “One can use film of naturally occurring events, acted events, or animation”[1976:125]. While the visual potential of events has been and continues to be largely exploited in ethnographic films, the possibilities of animation certainly deserve more exploration.
NOTES

1. An abridged version of this paper was presented—together with the film which it discus-
ses—at the symposium "Visual Research Strategies—Visual Anthropology in the 80s," at
the 12th International Congress of Anthropological and Ethnological Sciences, Zagreb,
24–31 July 1988. I would like to thank for their contributions: James Rosenstein for
correcting the English of this article; Jean Schwarz, ingenieur (E.R. 165 of the C.N.R.S.),
for the sonagrams made in the sound laboratory of the Ethnomusicology Department of
the Musée de l'Homme; Société COMANDA for the computer outprints; Christian
Lemzaouda (G.S. 74 of the C.N.R.S., Musée de l'Homme), for the photographic reproductions
of the sonagrams; Cinémathèque Française for the photogram Figure 11, extracted
from the film.

2. The original record notes are bilingual with texts in French and English. To make the text
available to the German-speaking part of Switzerland to which the people of the Muotatal
belong, I added for distribution in Switzerland a German language inlet.

3. A music book with conventional staff notations and written analyses is for many children
(and also for adults without training in score reading) too abstract, even if the book is
supplemented by an audiocassette illustrating it with sound recordings. Workshops with
musicians are a good way, but traditional musicians are not always available to come to
schools for workshops, and schools do not always have the money to remunerate them.

4. For an excellent use of hand graphs, see Reck [1977].

5. As this paper goes to press, our research group ordered a new model Sona-Graph,
permitting three minutes of spectral analyses in real time. I intend to explore its possi-
bilities in a new film.

6. Humor is not absent in the other three films of the series. For example, in the situations
where a photographer commands the family picture (The Wedding of Susanna and Josef), or a
drover holds the tail of a cow to be pulled on the steep path (Glattalp), in joking conversa-
tions (Yootzing and Yodelling), and in the surprising final disclosure of the editing devices
(Glattalp) [Zemp 1988b].

7. The English and the French versions had been shown to different audiences (ethno-
musicologists, music students, the larger public). The German version, just finished, had
not been presented yet to performers and other people of the Muotatal.

REFERENCES

Baumann, Max Peter

Betschart, Peter
1981 Der Muothaler Juuz. Bärgrüehlig (journal of the Swiss Yodel Conductor Associa-

Heider, Karl G.
1976 Ethnographic Film. Austin: University of Texas Press.

Reck, David

Sichardt, Wolfgang
1939 Der Alpenländische Jodler und der Ursprung des Jodelns. Berlin: Schriften zur Volks-
kunde und völkerkundlichen Musikwissenschaft, 2).

Young, Colin
Hague: Mouton.
Zemp, Hugo

FILMOGRAPHY

Zatzinger, Guy and Andrew Tracey
Zé, Moïse
Zemp, Hugo
1986a Yootzing and Yodelling. (French version: Youtser et yodler; German version: Juuzen und jodeln). Meudon, France: CNRS Audiovisuel.
1987 Head Voice, Chest Voice. (French version: Voix de tête, voix de poitrine; German version: Kopfstimme, Bruststimme). Meudon, France: CNRS Audiovisuel.