Timbre Variations as an Attribute of Naturalness in Clarinet Play

Snorre Farner[1], Richard Kronland-Martinet[2], Thierry Voinier[2], and Sølvi Ystad[2]

[1] Department of electronics and telecommunications, NTNU, O.S. Bragstads plass 2B, 7491 Trondheim, Norway
farner@iet.ntnu.no
[2] Laboratoire de Mécanique et d'Acoustique, CNRS, 31, chemin Joseph Aiguier, 13402 Marseille Cedex 20, France
{kronland, voinier, ystad}@lma.cnrs-mrs.fr

Abstract. A digital clarinet played by a human and timed by a metronome was used to record two playing control parameters, the breath control and the reed displacement, for 20 repeated performances. The regular behaviour of the parameters was extracted by averaging and the fluctuation was quantified by the standard deviation. It was concluded that the movement of the parameters seem to follow rules. When removing the fluctuations of the parameters by averaging over the repetitions, the result sounded less expressive, although it still seemed to be played by a human. The variation in timbre during the play, in particular within a note's duration, was observed and then fixed while the natural temporal envelope was kept. The result seemed unnatural, indicating that the variation of timbre is important for the naturalness.

1 Introduction

Naturalness is a term that is regularly used in the context of speech and music synthesis, although a clear definition is difficult to formulate. Ternström [1] has suggested a layered transport model of communication, be it musical or spoken, where in the present case a musical message at the first layer may be converted to a script of musical phrases in the second layer, then to a sequence of notes with certain attributes, further to gestures (e.g. movements of the bow of a violin) that are finally converted by the instrument to sound waves in a last layer. The sound waves are transmitted to the listener after being distorted by room acoustics and possibly by microphones and loudspeakers. Another important aspect is of course the layer of the listener's perception, but we will not be concerned with this here.

By considering naturalness in this way, one can suppose that a defect in any layer may cause the sound to be perceived as unnatural. As an example, in synthesis of musical instruments, lack of naturalness seems to occur either at the gesture level, for instance when a control interface cannot sufficiently capture the player's gestures, or at the instrument level, due to a poor instrument or radiation model. In general, for the purpose of music and speech synthesis,

R. Kronland-Martinet, T. Voinier, and S. Ystad (Eds.): CMMR 2005, LNCS 3902, pp. 45–53, 2006.

naturalness may be defined as the attribute that makes the listener think that the sounds are produced by a human. Rodet [2] has referred to this definition in the context of synthesis of the singing voice, and Nusbaum et al. [3] asked their subjects in a listening test whether the vowel they heard was produced by a human or by a computer, thus also adhering to this interpretation of the term naturalness.

An important issue concerning naturalness and sound synthesis is related to the fact that at present a computer cannot mimic the human speech and music performance in a convincing way, unless it copies the performance. Even when a natural sound is distorted during transmission, for instance by the telephone where higher frequencies are lost and noise is added, we do not doubt that the speaker is human. This example shows that an important part of the naturalness, as defined above, is contained in the control of the sound source rather than in the sound itself. We therefore search for cues that are important for rendering a music performance natural, in the sense that the listener thinks that it is performed by a human and not by a computer program.

A number of studies have already been conducted on performance rules defining the performer's deviations from musical scores. The rules may be divided in two main categories [4]: differentiation rules linked to duration, pitch, and intervals, and grouping rules linked to the way the performer gathers tones into melodical gestures, subphrases, and phrases. To our knowledge, no rules have been established that take into account variations in timbre during the notes. This is necessary for self-sustained instruments, such as the clarinet or the violin, as the sound is also controlled after the note onset. Hence, a new family of expressive parameters contributes to the performance and gives new degrees of freedom that act on the naturalness.

In the present study we take a closer look at the variation of such parameters and their relation with the naturalness in the clarinet case. More precisely we divide the naturalness at the gesture and instrument layers into a systematic and a random part. In fact, even a musician who controls his instrument to perfection cannot perfectly reproduce the exact same musical phrase. There are muscle vibrations, small variations in the player's lip position, phase variations of waves in the resonator and much more. It should be mentioned that the GERMS model [5] goes further and divides a performance into five principal expressive components: Generative rules, Emotional expression, Random variations, Motion (i.e. gesture), and Stylistic unexpectedness (hence the acronym). The systematic part should encompass all but the random variations of this model.

We address the systematic and random variations in the control parameters during each note and verify our hypothesis that these parameters follow rules in a similar way as do the variations between notes. We also link the variation in timbre, here represented by the spectral centroid, to the perception of naturalness.

Finally, "[Sound x]" refers to sound examples on the CMMR Internet site: http://www.lma.cnrs-mrs.fr/~cmmr2005/

2 Sound Preparations

In order to test this hypothesis, it is important to be able to measure the control parameters employed by the musician as well as parameters that the musician does not control, such as room acoustics. Although a real acoustical instrument would be preferred for studies on naturalness, an electronic instrument was chosen since it allows fine measurements of the control parameters as well as reinjection of these parameters into the synthesis model.

The physics of the clarinet has been studied for a long time, and simple relations describe quite realistically its sound production [6, 7, 8]. Without giving a detailed description of all quantities involved, the model can be summarized by three equations simulating the reed motion, the pipe resonances, and the nonlinear interaction between the reed motion and the pipe resonances. The reed motion is modelled as a spring with mass and damping:

$$\mu \frac{d^2y}{dt^2} + r\frac{dy}{dt} + ky(t) = p(t) - P, \tag{1}$$

where P is the mouth pressure, y the oscillating reed displacement from equilibrium, and p the oscillating pressure inside the mouthpiece. The resonances of the pipe may be described by its input impedance

$$\tilde{Z}(\omega) = jZ_c \tan(\omega L/c - j\alpha\sqrt{L}), \tag{2}$$

which relates p and u to the pipe length L. The Bernoulli equation may be used to couple the two equations together by the oscillating volume flow u of air through the mouthpiece:

$$u(t) = w(y(t) + H)\sqrt{\frac{2(P - p(t))}{\rho}}, \tag{3}$$

where H is the equilibrium height of the reed opening. This system of equations can be solved in many ways, but the main point is that the model provides the musician with three control parameters: the length of the pipe L, the blowing pressure P, and the equilibrium reed opening H.

This or a similary physical model is implemented in the Yamaha VL70-m virtual acoustic synthesizer [9] although we do not have access to the true contents and real-time implementation. The synthesizer may be piloted by various controllers, and we have here used the clarinet-like Yamaha WX5 MIDI controller. The fingering determines the note to play and thus the length L of the pipe, the blowing pressure P is captured as the MIDI breath-control parameter BC, and the reed opening H is represented by a MIDI "general control" parameter that we will call the (equilibrium) reed displacement RD. The latter is controlled by the lower lip against a non-oscillating reed. In addition, the reed displacement may also give a pitch bend, i.e. a slight variation of the pitch, but this was disconnected in the present study and is disregarded in the following. The musician's actions on the WX5 are transmitted by MIDI parameters to the VL70-m, which synthesizes the sound in real time based on a model similar to the one described

and depending on the variations in the fingering, the reed displacement, and the blowing pressure.

This setup gives the musician a realistic, albeit limited, control similar to a clarinet and thus offers some expressive control of the timbre of the sound. While the relation between the physical quantities P and H is not directly related to the MIDI parameters BC and RD, and not knowing exactly how they are treated in the synthesizer, we will only consider the MIDI parameters (scaled to range from zero to one) in the following.

A 20-second theme of the melancholic song "Plaisir d'amour" (Jean-Paul Martini 1760) [Sound 1] was played 20 times by an amateur clarinetist, who was asked to play all repetitions as equally as possible, with the same interpretation and expression. A metronome was used to facilitate comparison between the 20 repetitions. The MIDI data (breath control, reed displacement, note-onset timing with note value, and metronome timing) for each performance was saved to a text file, and the synthesized sound was recorded to a WAV file. The data were processed in Matlab.

3 Systematic and Random Variations

Figure 1 shows the MIDI parameters BC and RD for two of the 20 versions played [Sound 2, 3]. The vertical gray lines represent the note onset timings while the vertical dotted lines indicate the metronome ticks. The figure shows that BC and RD vary in a regular way, but with some variations from repetition to repetition. The systematic variation will survive an averaging over all the repetitions, while the random variations can be quantified by the standard deviation. A few precautions are necessary, however. First, all performances must be synchronized, which was facilitated by the use of the metronome. Second, the

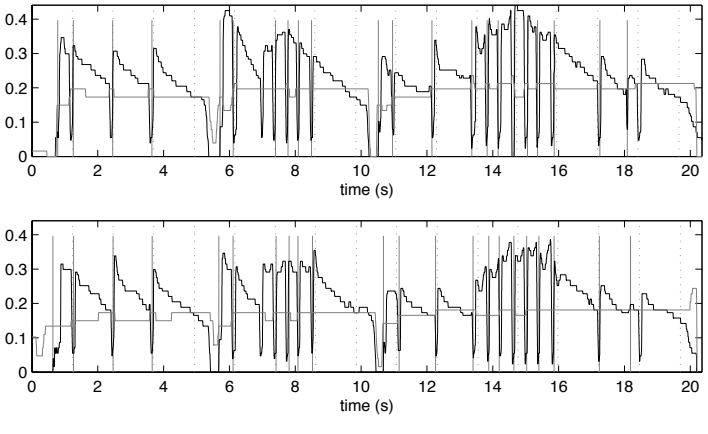

Fig. 1. Variations of the breath control BC (black lines) and the reed displacement RD (gray lines) for two different performances. The vertical lines mark note onset (gray) and metronome (dotted) timings.

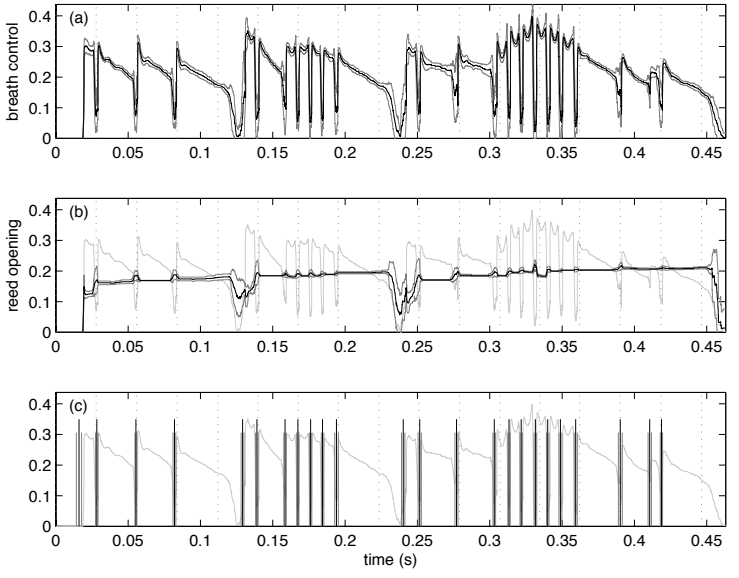

Fig. 2. The systematic and random changes of (a) the breath control BC and (b) the reed displacement RD represented by their normalized mean (black lines) and standard deviation (gray lines). The vertical lines in (c) show the note-on timings (black lines), their standard deviations (gray lines), and the metronome timings (dotted lines).

curves should be normalized note by note so that level variations between the repetitions do not contribute to the standard deviation.

The notes were separated by their note onset timings and shifted to the average timings. Normalization was performed individually for each note: The parameters were divided by the time average for each note, e.g. $\overline{BC}_i = \langle BC_i(t) \rangle$ for the blowing pressure. We ignored values below 25 % of the maximum BC and below 50 % of the maximum RD in order to avoid taking into account pauses and bumps between the notes. Then the normalized curve was averaged over all repetitions $i = 1, \ldots, 20$ and finally remultiplied by the mean of all $\overline{BC}_i$ for this note. The standard deviation was calculated in the same way by normalizing by $\overline{BC}_i$. In this way, the final normalized standard deviation should only include the random variations within each note, i.e. the variations in curve form and not variations in the general level of the notes.

The mean and standard deviation of BC and RD are presented in Figure 2a and b, respectively, while the standard deviation of the note-on timings are shown as vertical gray lines around the solid mean lines in Figure 2c.

The mean of BC shows that there is a systematic variation of the breath control during the play. The long notes have a peak at the start and decrease gradually towards their end. The breath control falls to a minimum for a short moment before the next note is attacked. In the more rapid parts in the middle of the second and third phrases, however, the movement of the notes shows a different shape, especially the short ones: there is also a peak at the end of the

note. Additionally, the crescendo in the beginning of the last phrase is manifested by smaller slopes in the long notes and a higher blowing pressure in the middle of the phrase. These characteristics indicate that a set of playing rules may be established. Although a part of the movement is necessary for a sound to be produced, the musician has some liberty to express his intention through manipulation of the control parameter.

This is also the case for the mean of the reed displacement RD, though this parameter to a great degree is kept constant during the notes. The constant reed opening is coherent with the playing technique that is most commonly used among classical musicians. Although the playing style will depend on the construction of the clarinet, changing the reed opening in the model presented in Section 2 makes the brightness of the sound change [10]. In classical music, it is in general desired that the brightness does not vary from note to note, and on the described model, the reed opening may be used to compensate changes in brightness due to variations in the note value and in breath control.

Another characteristic of the mean variation of RD is that there are two deep bumps at the transition between each musical phrase. At these points the player took his breath and thus relaxed the pressure on the reed for a moment. Breathing is a natural part of a clarinet performance and may contribute to the naturalness. The sound of breathing is not included in the synthesis, but it seems evident for the listener where the player breathed. The pause might be sufficient for the perception of natural breathing, but it is possible that the bumps in the reed displacement affect the termination of the previous note and the attack of the following, and thus contribute to the perception of naturalness.

It is interesting to note that playing the averaged parameters by the VL70-m synthesizer gives a result that seems to lack some expressiveness [Sound 4]. However, it may still be considered natural in the sense that a human seems to have played it, although maybe less motivated.

The standard deviation of RD is very small in this study. One reason for this is the steady reed opening, but it is perhaps also due to the discretization induced by the MIDI protocol, as was suggested in Figure 1. Whether the subtle nuances that would result from a better MIDI resolution would be audible has to be investigated in a future work.

4 Timbre Variations

Timbre is defined as the perceptual attribute that distinguishes two tones of equal pitch, loudness and duration [11]. Furthermore, the possibility to *vary* the timbre continuously during the play seems to be important for the musician wishing to interpret the music expressively. It also seems that such continuous variations of the timbre contribute to the naturalness of a performance.

In the following, we look at the variation of the brightness [12] of the sound, which is a commonly used timbre descriptor. It is defined by

$$SCG = \frac{\int_0^{f_c} |\hat{S}(f)| f df}{\int_0^{f_c} |\hat{S}(f)| df} \tag{4}$$

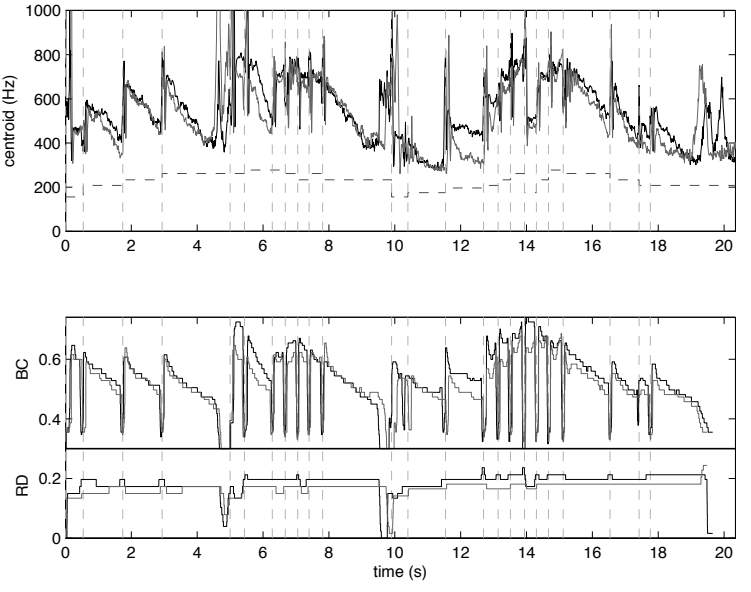

Fig. 3. The spectral centroid of the sound resulting from the two repetitions in Figure 1 (in black and gray) together with the note frequency (dashed line). The corresponding breath control BC and reed displacement RD curves are shown below, and dashed vertical lines show note onsets.

where f and $|\hat{S}(f)|$ respectively represent the frequency and the magnitude of the Fourier transform of the signal $s(t)$. A continuous curve of the centroid frequency was obtained by taking the short-term Fourier transform of 50 % overlapping signal frames of 23 ms, weighted by a Hanning window. The cut-off frequency f_c was introduced because the VL70-m synthesis added noise to simulate blowing noise, which was more pronounced for small blowing pressures. The noise disturbed the calculations of the centroid and was removed by setting $f_c = 2500\,\text{Hz}$. Figure 3 shows the spectral centroid frequency for the two repetitions in Figure 1.

Without the noise, the correlation between the blowing pressure and the centroid is obvious: The sound is brighter in the beginning of the long notes, and darkens as the blowing pressure decreases. This is in accordance with the "Worman rule" that the pressure wave is nearly sinusoidal for low excitation energy (for weak blowing pressure), and that the higher harmonics become more important as the player blows harder [13, 14]. The correlation with the reed displacement RD was ignored in the present study as this parameter was mostly constant during the notes.

The decrease in the centroid frequency was about 300 Hz for the long notes, accompanied by a decrease in energy (not shown here). We have verified by listening that the change in the spectrum from the beginning to the end of these notes was highly audible, even when compensating for the energy decrease. This was done by extracting frames at several points from the beginning to the end of a long note, removing the blowing noise by low-pass filtering, normalizing

them to approximately the same loudness, and applying a pitch-synchronous lengthening to make listening possible [Sound 5].

To show the effect of the timbre on the naturalness, we can force the timbre to be constant during the play. This may be done by freezing BC and RD of one of the repetitions to their mean values and feeding them together with the old note-change parameters to the synthesizer. The result sounds static and unnatural, much because also the intensity of this static performance was fixed [Sound 6]. We therefore forced the intensity to vary as for the original performance by multiplying the signal by the envelope of the original. Although it might be accepted that this was played by a human musician, the timbre did not vary in a natural way [Sound 7]. We take this as an indication that variation of the timbre is important for the perception of naturalness.

5 Conclusions and Further Work

The analysis of the 20 repetitions of the 20 s music performance shows that a regular pattern of the two control parameters, breath control (BC) and reed displacement (RD), can be extracted by averaging the performances, while the variations between repetitions, considered the random part, may be quantified by the standard deviation.

For the regular movement of the breath control, qualitative differences was found between notes in the sequences of short notes and the calmer parts of the melody. This suggests that expressive rules may be established for the movement of BC. The reed displacement was found to vary little during each note, at least within the MIDI resolution. This was coherent with the fact that the player had been taught to play with a steady reed.

Because the parameters were normalized, the standard deviation mainly showed the difference in *shape* between the curves of each note. When reconstructing a performance from the averaged parameters, the result seemed to lack some expressiveness compared to any of the 20 real performances. This suggests that the parameter fluctuations quantified by the standard deviation contribute to the expressiveness.

We have verified that variations in the breath control BC induce variations in the timbre, at this stage only quantified by the spectral centroid. When fixing the timbre, but keeping the temporal envelope of the signal, the result becomes unnatural. This indicates that the variation of the timbre is important for the perceived naturalness of the performance.

To elaborate on these preliminary results, we foresee to

- apply a more recent synthesis method [15] that is considered more realistic and where more information on the timbre is available,
- use a professional musician and a larger repertoire of playing styles,
- effectuate a more detailed study of the rules related to the movement of the parameters,
- employ other measures of the timbre, and
- perform listening tests to determine the importance of variation of timbre and intensity as well as timing to the perception of naturalness.

Finally, an interesting question is whether such rules depend on musicians and on playing styles. If so is the case, recognition of musician and playing style from MIDI data could be considered, and it should make it possible to use rules obtained from analysis of MIDI performances to add expression and naturalness to simple score data.

Acknowledgements

This work has been partly supported by the Research Council of Norway.

References

1. Sten Ternström, "Session on naturalness in synthesized speech and music," in *143rd ASA meeting, Pittsburgh*, June 2002.
2. Xavier Rodet, "Synthesis and processing of the singing voice," in *IEEE Benelux Workshop on Model Based Processing and Coding of Audio (MPCA)*, Leuven, Belgium, Nov. 2002, pp. 99–108.
3. Howard C. Nusbaum, Alexander L. Francis, and Anne S. Henly, "Measuring the naturalness of synthetic speech," *International Journal of Speech Technology*, vol. 1, no. 1, pp. 7–19, 1995.
4. Johan Sundberg, "Grouping and differentiation. Two main principles in the performance of music," in *Integrated Human Brain Science: Theory, Method, Application (Music)*, T. Nakada, Ed., pp. 299–314. Elsevier Science, 2000.
5. Patrik N. Juslin, "Five facets of musical expression: a psychologist's perspective on music performance," *Psychology of Music*, vol. 31, no. 3, pp. 273–302, 2003.
6. R. T. Schumacher, "Self-sustained oscillation of the clarinet: an integral equation approach," *Acoustica*, vol. 40, pp. 298–309, 1978.
7. J. Kergomard, "Ch. 6. Elementary considerations on reed-instruments oscillations," in *Mechanics of Musical Instruments, Lecture notes CISM*, A. Hirschberg, J. Kergomard, and G. Weinreich, Eds., pp. 229–290. Springer Verlag, Wien/New York, 1995.
8. J.-P. Dalmont, J. Gilbert, and S. Ollivier, "Nonlinear characteristics of single-reed instruments: Quasistatic volume flow and reed opening measurements," *J. Acoust. Soc. Am.*, vol. 114, pp. 2253–2262, Oct. 2003.
9. R. Rideout, "Yamaha VL1 virtual acoustic synthesizer," *Keyboard Magazine*, vol. 20, pp. 104–118, June 1994.
10. Robin Thomas Helland, "Synthesis models as a tool for timbre studies," Master thesis, Norwegian University of Science and Technology, 2004.
11. ANSI. American National Standard — Psychoacoustical Terminology, (American National Standards Institute, Inc., New York).
12. James W. Beauchamp, "Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones," *Journal of the Audio Engineering Society*, vol. 30, no. 6, pp. 396–406, 1982.
13. W. E. Worman, *Self-sustained nonlinear oscillations of medium amplitude in clarinet-like systems*, Ph.D. thesis, Case Western Reserve University, 1971, Ann Arbor University Microfilms (ref. 71-22869).
14. Arthur H. Benade, *Fundamentals of musical acoustics*, Dover Publication, New York, 2 edition, 1990, 1st ed. published by Oxford University Press in 1976.
15. Ph. Guillemain, J. Kergomard, and Th. Voinier, "Real-time synthesis models of wind instruments based on physical models," in *Proc. of the Stockholm Music Acoustics Conference (SMAC)*, Sweden, Aug. 2003, pp. 389–392, Stockholm.