# E-LEARNING SOFTWARE FOR IMPROVING STUDENT'S MUSIC PERFORMANCE USING COMPARISONS

Delgado, M., Fajardo, W. and Molina-Solana, M.
*Universidad de Granada. Dept. Computer Science and Artificial Intelligence, Granada. Spain*

## ABSTRACT

In the last decades there have been several attempts to use computers in Music Education. New pedagogical trends encourage incorporating technology tools in the process of learning music. Between them, those systems based on Artificial Intelligence are the most promising ones, as they can derive new information from the inputs and visualize them in several meaningful ways. This paper presents an application of machine learning to music performance which is able to discover the similarities and differences between a given performance and those from other musicians. Such a system would help students to better learning how to perform a certain piece of music, allowing them to compare with other students or master performers.

## KEYWORDS

E-Learning, Music education, Music Performance, Machine Learning.

## 1. INTRODUCTION

Student musicians spend most of their time practising. These countless hours of practising help them learn to interpret a piece of music as the composer imagine it, but also developing their own style -one that is unique to each of them. Said in other words, what makes a piece of music come alive is also what distinguishes great artists from any other.

In this work, we propose a piece of software and a machine learning algorithm which are able to classify performances according to their similarity to those of famous musicians. Such a system would help students to understand the musical resources some of the greatest performers use and how to imitate them.

We focus on the task of characterizing performers from their playing style using descriptors that are automatically extracted from commercial audio recordings by means of state-of-the-art feature extraction tools. This learning process is done by employing the system described in Molina-Solana *et al.* (Molina-Solana, M. et al., 2010). This approach is quite different from those in the literature, in which a heavy human intervention is needed in order to manually annotate the music. That is, however, what makes it feasible for being employed in an automated system such as the one we present here. The devised software requires no human intervention, being the whole process done in an automatic way.

Software tools aimed to help students to better perform have mainly the drawback that they are limited to a range of previously analysed songs. That is not the case of our system as far as the learning model we have chosen, is able to capture how the performer plays, regardless the piece being performed. This way, the student is assessed of how to play a work even though the expert musician did never play it.

No special hardware (e.g. MIDI instruments) is required to appraise the student's unique way of performing. Only an audio recording of the student performing with their own instrument is required. Thus, the performance is no biased in any sense: students play with their own instrument and there is no need of them being aware of them being recorded. In fact, it is relatively easy to count with students' recordings nowadays because many teachers encourage their students to record themselves when playing.

This paper is organized as follows. Firstly, in Section 2 we provide some background about both Music Performance and Music Education. In section 3, we briefly explain the model we use to represent the information about performances, and how that information is extracted from audio files. Section 4 presents the devised software and how it can be used. Finally, Section 5 points out some conclusions and further developments.

## 2. BACKGROUND

This section describes what Music Performance is, and how computers can be used for Music Education. Former works in those areas are also revised to provide the reader a proper framework for the rest of the paper.

### 2.1 Music Performance

The literal execution of a musical score is frequently judge as being significantly less interesting than a performance of that piece by even a moderately skilled musician. The reason is straightforward: musicians do not produce literal renditions of the score. That is, they do not play a piece of music mechanically, with constant tempo or loudness, exactly as written in the printed music score. Rather, skilled performers slow down at some places, speed up at others and stress certain notes or passages. Tempo variations and loudness variations are the most important parameters available to a performer, being also the main source of expression in music (Palmer, C., 1996). The way these parameters 'should be' varied is not precisely specified in the printed score; so that, it is the performer the one in charge of using them properly.

According to Widmer and Goebl (Widmer, G. and Goebl, W., 2004), expressive music performance can be defined as the deliberate shaping of the music by the performer, in the moment of playing, by means of continuous variations of parameters such as timing, loudness or articulation. Changes in tempo (timing) are modifications of the regular grid of beats that defines time in a score. Changes in loudness (dynamics) are modifications of the intensity of notes with respect to the others and to the general energy of the fragment in consideration. Articulation consists in varying the gap between contiguous notes by, for instance, making the first one shorter or overlapping it with the next.

Music performance is a complex activity involving physical, acoustic, psychological, social and artistic issues. At the same time, it is also a deeply human activity, relating to emotional as well as cognitive and artistic categories. Furthermore, it is dramatically affected by performer's mood and physical condition.

As said before, research in music performance has a multidisciplinary character, with studies that range from understanding expressive behaviour to modelling aspects of renditions in a formal quantitative and predictive way. Historically, research in expressive music performance has focused on finding general principles underlying the types of expressive 'deviations' from the musical score (e.g., in terms of timing, dynamics and phrasing) that are a sign of expressive interpretation. Works by De Poli (Poli, G. D., 2004), Widmer and Goebl (Widmer, G. and Goebl, W., 2004), and Delgado *et al.* (Delgado, M. et al., 2011) contain overviews on expressive performance modelling. The reader should refer to them for further information.

One of the issues in this area is the representation of the way certain performers play by just analysing some of their renditions -studies into the individual style of famous musicians. That information would enable us to identify performers by just listening to their renditions (Molina-Solana, M. et al., 2010; Saunders, C. et al., 2008; Stamatatos, E. and Widmer, G., 2005; Widmer, G. et al., 2003). These studies are difficult because the same professional musician can perform the same score in very different ways. Among the methods for the recognition of music performers and their style, the most relevant are the fitting of performance parameters in rule-based performance models, and the application of machine learning methods for the identification of performing style of musicians. Recent results of specialized experiments show surprising artist recognition rates (for instance, see those from Saunders *et al.* (Stamatatos, E. and Widmer, G., 2005) or Molina-Solana *et al.* (Molina-Solana, M. et al., 2010)).

### 2.2 Computers in Music Education

According to Brown (Brown, A.R., 2007), there are three main roles in which a computer can take part in music education: they can act as a tool, a medium or a musical instrument. A tool for recording, editing, analyzing and sequencing sounds; a medium for music storage, indexing and distribution; and a musical instrument for synthesizing music in real time.

In the last decades there have been several attempts to use computers in Music Education. The field is highly interdisciplinary, involving contributions from disciplines such as Music, Education, Artificial Intelligence, Psychology, Linguistics, Human Computer Interaction and many others. Advances in research,

as well as new software tools for the analysis of data, open up a new area in the field of music education, as stated by Friberg and Battel (Friberg, A. and Battel, G. U., 2002).

Artificial Intelligence in Music Education is also a very diverse field, being the work by Holland (Holland, S., 2000) still an interesting review in the topic. Attempts in this area can be classified into four categories, according to Brandao *et al.* (Brandao et al., 1999), which are intended to: teach fundamentals of music, teach musical composition skills, perform analysis of music, and teach musical performance skills. In this paper, we propose an example within the scope of the last group.

However, music performance is not a well-defined domain. There are no clear goals, correct answers or an only way of doing things. That is probably why it is a tough domain researchers are not willing to face. Persson *et al.* (Persson, R. S. et al., 1996) pointed it out when they stated that relatively little time is dedicated to interpretative and emotional aspects of performance in comparison with that dedicated to learning notation -a domain that can be more easily represented.

So, why using technology for such a task? Despite computers will hardly replace music teachers, they can complement them. The benefits of individualized instruction, assessment and motivation can be used to supplement the learning. Using technology, students can work at their own pace, focusing on those aspects they need to improve. Because the learning is individual, there is no peer competition that, in many cases, is counterproductive. Evaluation of responses can be totally accurate and impartial. These possibilities have been tested with promising results by Friberg and Battel (Friberg, A. and Battel, G. U., 2002).

New pedagogical trends suggest recording students' performances. This way, they can hear how they play and analyse critically their work. Listening to both professional and amateur recordings is also a valuable teaching opportunity for music students. On the one hand, this allow them to compare their own current performance to those of professional musicians, being a motivational way of helping students to set goals for practising. On the other hand, students can identify improvements with respect to previous recordings and be aware of their progression and achievements during the time.

Beside the role recordings play to identify errors, the use of recordings can also be useful for going further and focus on specific musical elements or sections, providing a deeper understanding of the performance. Furthermore, it gives students the possibility of record themselves in other places different than the classroom, releasing them of the pressure of a classroom situation and allowing them to better perform.

Once recordings are available, we can also take advantage of them by means of computational tools, like the one we are presenting in this paper. These tools might be able to point out the tempo of the work, to discover irregularities in the rhythm, or to stress some musical parameters. Of special interest is the possibility of representing the sound in a new meaningful way, like in the works by Lagner and Goebl (Langner, J. and Goebl, W., 2003) or Saap (Sapp, C., 2007). In our case, we are able to offer a list of performers that played in a similar way of the input recording.

The analytical comparison between a natural performance and performances with particular expressive intentions also seems to possess a potential for music pedagogy. One of these tools is *Director Musices* (Friberg, A. et al., 2000) which is able to play a song applying some performance rules. This way, certain acoustic parameters can be exaggerated so that anybody, regardless of their musical training, can detect the differences and concentrate on particular aspects of the performance.

## 3. LEARNING ALGORITHM

Our approach for dealing with the characterization of performers is based on the acquisition of trend models, which represent each particular performance. As said before, we employ the learning algorithm described in Molina-Solana *et al.* (Molina-Solana, M. et al., 2010). It is a quick and automatic method for collecting information about the performer of a musical piece from its audio. It obtains an approximate score with information about pitch and energy. Because the method is automatic, there is no manual annotation and a confident musical analysis is not performed.

Despite those apparent drawbacks, the alternative the trend model presents consist in taking a more global perspective of a performer by trying to capture his essential (and recurrent) expressive decisions. This method does not focus in concrete cues, but in the general tendency. Because only an approximated score derived from the audio is available, a melodic contour segmentation is employed as a way to capture the

musical structure of the piece. This segmentation allows an analysis in terms of melodic intervals, i.e. precision on pitch detection is not critical.

Specifically, a trend model characterizes, for a specific audio descriptor, the relationships a given performer is establishing among groups of neighbour musical events. A qualitative analysis of the variations of the audio descriptors is performed with a local perspective. Two trend models will be used: energy and duration. The trend model for the energy descriptor relates, qualitatively, the energy variation for a given set of consecutive notes, and it is related to dynamics. On the other hand, the trend model for duration indicates, also qualitatively, how note durations change for note sequences. Duration is related to articulation and timing.

Given an input musical recording of a piece, the trend analysis is performed by aggregating the qualitative variations on their small melody segments. Thus, in advance of building trend models, input streams are broken down into segments of three-note long. As most automatic melody segmentation approaches do, note grouping is performed according to a human perception model. At the training stage, the goal of our system is to characterize performers by extracting expressive features and constructing trend models. Next, at the identification stage, the system analyses the input performance and looks for the most similar previously learned model.

## 3.1 Data Representation

As said, the first step consists in extracting audio features. So far, we only consider fundamental frequency and energy, as these are the main low-level audio features related to melody. These features are then used to identify note boundaries and to generate melodic segments. This module provides a vector with instantaneous fundamental frequency and energy values calculated every 0.01 seconds.

After some post-processing for deleting noise, a smooth vector of pitches is obtained. By knowing on which samples pitches are changing, a note-by-note segmentation of the whole recording is performed. For each note, its pitch, duration and energy are collected.

It is assumed that there might be some errors in this automatic segmentation, given the heterogeneity of recording conditions. The trendmodel algorithm proposes a more abstract representation (but still close to the melody) than the real notes for dealing with this problem. That is, instead of focusing on the absolute notes, we are interested in modelling the melodic surface.

To do that, a simple contour model that identifies some melodic patterns from the melody is used. These patterns are three-note long and are related with the direction of the two intervals that exist in them.

## 3.2 Classification

A nearest neighbour classifier is used to generate the list of similar performers for each input recording. Trend models acquired in the training stage, as described in the previous section, are used as class patterns. When the student's recording is presented to the system, its trend model is created and compared with the previously acquired ones. The system outputs a ranked list of performer candidates where distances determine the order, with 1 being the most likely performer relative to the results of the training phase.

The distance $d_{ij}$ between two performances $i$ and $j$, is defined as the weighted sum of distances between respective contour patterns. Weights have been introduced for balancing the importance of the patterns with respect to the number of times they appear. Frequent patterns are considered more informative due to the fact that they come from more representative samples. When several audio descriptors are considered, the individual corresponding distances are aggregated. Readers should refer to the original paper describing this algorithm (Molina-Solana, M. et al., 2010) in order to find more details and the exact equations.

## 3.3 Model Performance

The feasibility and accuracy of the system have already been tested by using several works from Sonatas and Partitas by J.S. Bach (Molina-Solana, M. et al., 2010). In those experiments, the model was employed to identify several violinists by using some already labelled performances. The results were quite promising: much better than those from a random chance classifier and probably better that those achievable by a human

expert. It was also demonstrated that the information that trend models contain can be used to derive meaningful musicological data.

The system has only been tested with violin recordings, but the underlying machine learning model (trendmodel algorithm) is flexible enough to allow the employment of other monophonic instruments. The application itself is capable of dealing with any instrument or song as long as the information it receives is properly formatted.

## 4. SOFTWARE

We have implemented in MATLAB all the tools for extracting the information from musical signals, representing those data, and using it for comparing: the whole trend-model algorithm and an interface to use all the developed tools.

The software is very user-friendly. The procedure is as follow: 1) the user/student selects the file with their performance; 2) selects the files to compare with; 3) waits for the audio analysis at the learning stage; 4) launches the comparator.

The devised interface (see Figure 1) enables the user to load a set of files to be used as comparing performances. On the left, we find a list of all the audio files with performances that we have in the folder. From those, we can select only those we are interested in being used. On the right, we select the audio file containng the student's performance we want to compare with the others. The developed software uses audio files in wav format as input.
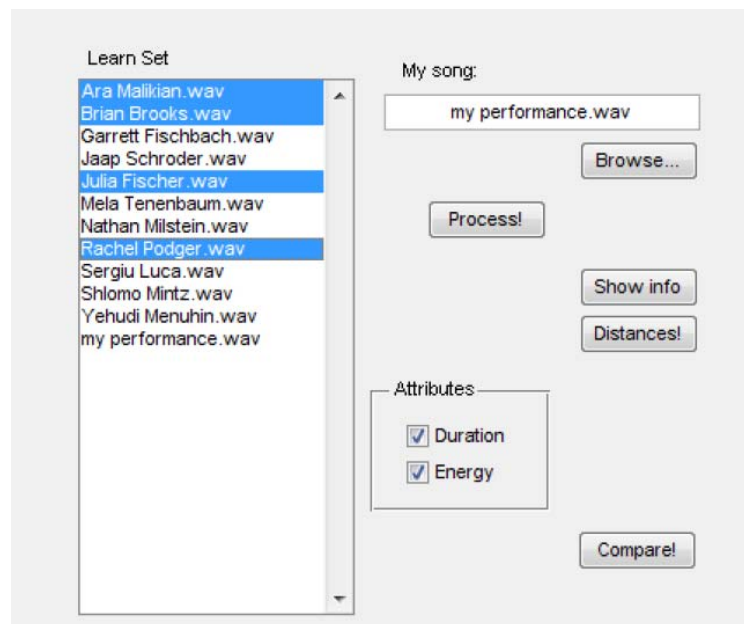


Figure 1. Main window of the devised interface of the system

Once we have selected the files we want to use, the system needs to extract information from them, and represent them by means of trend models. That is done when the 'Process' button is pressed. The interface offers the student the possibility to choose which (or all) attributes they want to use when comparing performances ('Attributes' section in the interface). Currently, only two possibilities are allowed as they are the attributes used in the learning stage: duration of notes and their energy. Those two attributes are the main ones in music performance, as previously explained in Section 2.1. Thus, students can compare themselves in terms of timing, in terms of energy, or in both together. Results may differ depending on the selection. However, any other musical attribute can be used as long as they are extracted from the audio and learned in the learning stage.

'Distances' shows a distance matrix comparing the distances between each pair of performances (including the one by the student). An example can be seen in Figure 2. A 2D-cluster representation of the

performances is also shown (Figure 3). This representation is done by comparing the distances between each pair of performances (in a high dimension space) and projecting them to a plain. For instance, here we can see that the student's performance is pretty similar to that of Jaap Schroder (always according to our described method for representing performances), and quite different from Garret Fischbach's.
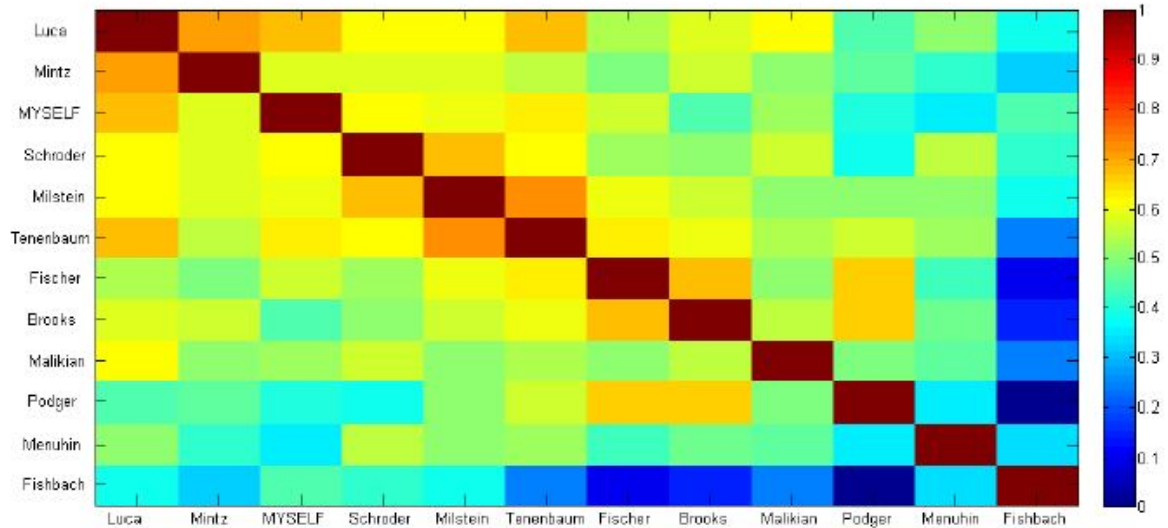


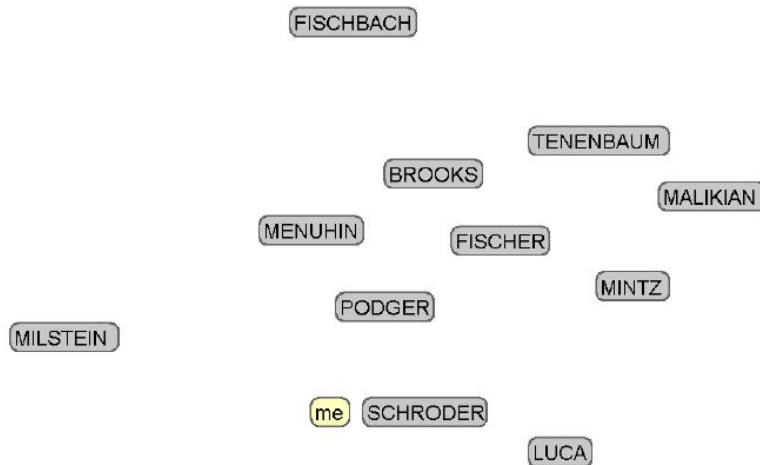Figure 2. Normalized similarity matrix between performances



Figure 3. Cluster representation of the performances

This cluster representation gives an idea of how the student plays in comparison with others. However, it cannot tell where the differences are nor the changes to be made to their style in order get closer to a certain performer. 'Show info' offers a representation of trend models. They are represented as histograms. Each bar indicates how a performer tends to play for a certain musical pattern. More details can be found in the paper where the algorithm is described (Molina-Solana, M. et al., 2010).

Finally, we can obtain a list of sorted performers ('Compare' button), with the first one being the most similar to the student one, and the last being the most different. The software offers the result as an image (see Figure 4). In black, we find the most similar performer, whereas in white is the furthest one.
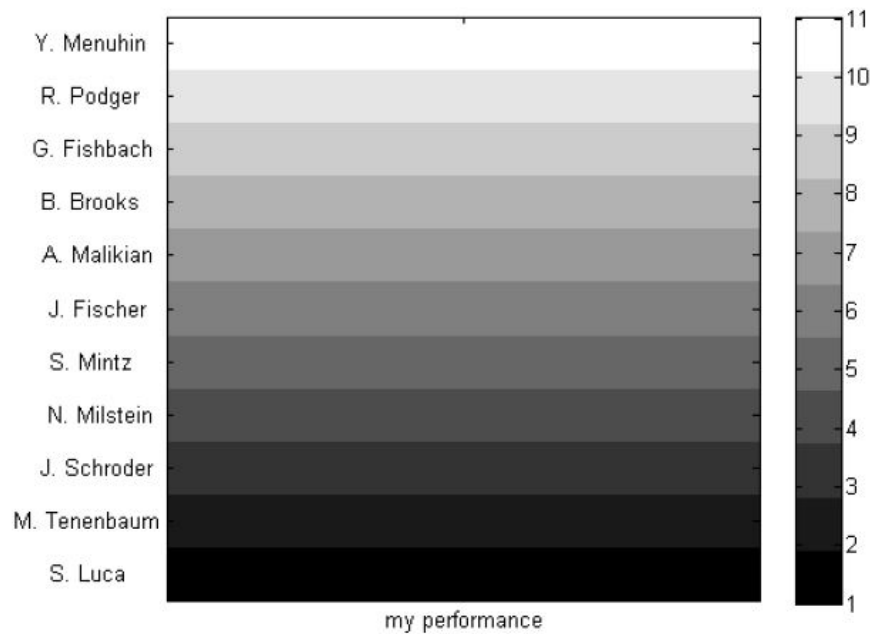
Figure 4. Ranking of similar performers

## 5. CONCLUSION

In this paper, we have presented an application of machine learning to music performance. The software we have developed is able to discover and show the similarities and differences between a given performance and those from other musicians. In our opinion, such a system would be of great interest for music students, as it can help them to better perform a piece of music. It shows them the similarities and differences with those renditions of famous performers.

Technology applied to music education is a very promising area. New pedagogical tendencies suggest recording students' performances. Our system uses precisely those recordings to proceed, not needing any other special hardware to collect the information. Doing so, we get another advantage: all the data is gathered in a transparent way, not interfering with the performance. We have employed and algorithm already in the literature which matches well with our constraints.

Research on music performance can point out expressive resources that traditionally have been hiding in musicians' skill and musical intuition. When explicitly formulated, these resources will give the user the possibility to play music with different expressive colouring. Even more, they would allow a computer to do so. Not in vain, the art of performing music is not well-defined and it is the result of several years of training; a knowledge that is very difficult to be appraised by a computer. Thus, studies in music performance are of great value in our time.

## REFERENCES

Brandao, M. et al., 1999. Computers in Music Education. *Proceedings of the AISB'99 Symposium on Musical Creativity.*

Brown, A.R., 2007. *Computers in Music Education: Amplifying Musicality*. Routledge.

Delgado, M. et al., 2011. A state of the art on computational music performance, *Expert Systems with Applications*, 38(1):155-160.

Friberg, A. and Battel, G. U., 2002. *The Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning, chapter Structural Communication*, pages 199-218. Oxford University Press, New York.

Friberg, A. et al., 2000. Generating musical performances with Director Musices. *Computer Music Journal*, 24(3):23–29.

Holland, S., 2000. Artificial Intelligence in Music Education: a critical review. *Readings in Music and Artficial Intelligence, Contemporary Music Studies*, 20:239–274.

Langner, J. and Goebl, W., 2003. Visualizing expressive performance in tempo-loudness space. *Computer Music Journal*, 27(4):69–83.

Molina-Solana, M. et al., 2010. Identifying violin performers by their expressive trends. *Intelligent Data Analysis,* 14(5):555-571.

Palmer, C., 1996. Anatomy of a performance: Sources of musical expression. *Music Perception,* 13(3):433–453.

Persson, R. S. et al., 1996. *Fostering the growth of high ability: European perspectives,* chapter Motivational and influential components of musical performance: A qualitative analysis, pages 287–302. Ablex, Norwood, NJ.

Poli, G. D., 2004. Methodologies for expressiveness modelling of and for music performance. *Journal of New Music Research,* 33(3):189–202.

Sapp, C., 2007. Comparative analysis of multiple musical performances. *Proceedings of 8th International Conference on Music Information Retrieval (ISMIR2007)*, pages 497–500, Vienna, Austria.

Saunders, C. et al., 2008. Using string kernels to identify famous performers from their playing style. *Intelligent Data Analysis*, 12(4):425–440.

Stamatatos, E. and Widmer, G., 2005. Automatic identification of music performers with learning ensembles. *Artificial Intelligence,* 165(1):37–56.

Widmer, G. and Goebl, W., 2004. Computational models of expressive music performance: The state of the art. *Journal of New Music Research,* 33(3):203–216.

Widmer, G. et al., 2003. In search of the Horowitz factor. *AI Magazine*, 24(3):111–130.