Morphing Dynamical Sound Models

Axel Röbel
Sekr. EN-8, Technical University of Berlin
Einsteinufer 17, 10587 Berlin, Germany
Tel: +49-30-314 25699, FAX: +49-30-314 21143,
email: roebel@kgw.tu-berlin.de

Abstract

In the following article we investigate a new approach to sound morphing that is strictly working in the time domain. It is based upon a method of modeling and synthesizing natural sounds with neural networks that has been introduced recently and is called dynamic or attractor modeling. By means of basic synthetic signals we investigate the fundamental properties of the proposed morphing scheme. Using two real world sound signals obtained from a saxophone we demonstrate the potential of the method when applied to complex natural sounds.

1 Introduction

Modeling time series is one of the important applications of the theory of non-linear dynamical systems in the field of neural networks. The basic mathematical foundation for this application, the reconstruction theorem, has been formulated first in 1980 [14; 13]. The first application of the reconstruction theorem in the field of neural networks is generally contributed to Lapedes and Farber for their investigation of the prediction of chaotic time series [2]. Later it has been discovered that neural networks might not only predict chaotic time series, but, may even synthesize time series with characteristics that are close to those of the training signal [6; 9]. For this application the term dynamic modeling has been introduced. One main problem with dynamic modeling is that the stability of the models is not guaranteed. Besides training a large number of models and selecting a stable one the use of short term recurrent training is generally considered to improve model stability [5; 1].

Dynamic modeling is not restricted to chaotic dynamics, but, may be used for any kind of stable dynamical system. In recent investigations it has been shown, that the dynamical models of natural sound signals may be used to resynthesize sound signals with high quality [10]. Up to now, the possibilities to control the resynthesized signals have been confined to the time evolution of the dynamics (duration of attack, hold and decay time). This is rather limited compared to standard sound synthesis methods, as for example additive synthesis. Due to the complicated relations between the network parameters and the attractor of the model it appears to be impossible to find any sensible algorithm that allows a user to change the model characteristics by directly altering the model parameters. Therefore, in the present article we follow another approach to allow for signal modification, which we will call attractor morphing. This approach is motivated by a recent investigation into the homotopic mixture of vector fields [3], which demonstrates the generation of new system dynamics by means of a homotopic mixture of chaotic progenitor systems. In the following article we will show, that the same homotopic mixing algorithm may be used to generate new sound signals from the dynamic models of two progenitor sounds. The results are particularly satisfactory because for sufficiently related progenitor attractors, the morphed signals exhibit an important feature of natural sound transitions of harmonic sounds that is the synchronized shifting of pairwise related partials [15].

The following article is organized as follows. In section 2 we shortly explain some fundamental aspects of the attractor or dynamic modeling of sounds. In section 3 we review the ideas of reconstructing state space from time series. In section 4 we introduce the new morphing algorithm and in section 5 we present and analyze the morphing results for a number of artificial sound signals. In section 6 we apply our algorithm to two real world saxophone signals. Section 7 concludes with a short summary of the results.

2 Attractor modeling of sounds

Traditional musical instruments belong to the class of dissipative, nonlinear mechanical systems. Due to the nonlinearity involved in the system dynamics and due to the complex relations involved, this type of system is generally difficult to analyze. In principle, however, the evolution of the system may be described in an n-dimensional state space S by means of a vector field, or if we confine ourselves to discrete time signals, by means of a mapping $f(.)$

$$\vec{z}_{k+1} = f(\vec{z_k}) \quad \vec{z} \in S \quad , \tag{1}$$

which connects the system state $\vec{z}$ at time k, with the system state at the next time step. For a stable system $f(.)$ and for large $k \to \infty$ the state $\vec{z_k}$ will be confined to a bounded and closed subset $L \subset S$ of the state space, which, generally, will have much smaller dimension then S. $L(\vec{z_0})$ is called the limit set of the initial state $\vec{z_0}$. If there exists an open set U such that for all $\vec{z} \in U$ the limit set $L(\vec{z}) = L' \subset U$ then the set L' is called an attractor with the basin of attraction U [11]. Due to this definition an attractor is stable in the sense that small disturbances do not push the system away from it. An attractor may be as simple as a point in state space or as complex as a fractal set if the system dynamics are chaotic. If the dynamical system, or the musical instrument, is observed by means of an output signal $s(n)$, the characteristics of this signal, or the sound, are closely related to the topology of the attractor [4].

In the context of real world sound signals the underlying system is always stable and, therefore, a stationary sound signal is always related to an attractor. The type of the attractor depends on the musical instrument and its excitation. In most cases sounds obtained from chaotic attractors are considered noise and, therefore, are not used by classical musicians. Therefore, the use of musical instruments is often confined to periodic or quasi periodic attractors. However, musical instrument are not used with a stationary excitation. For slowly varying dynamics this situation can be described by a system undergoing a parameter variation and, therefore, following a sequence of attractors [12; 10].

3 Reconstructing attractors

Assume an n-dimensional discrete time dynamical system $f(.)$ evolving on an attractor A with dimension $d < n$. The system state $\vec{z}$ is observed through a sequence of measurements $h(\vec{z})$, resulting in a time series of measurements $y_k = h(\vec{z}_k)$. Under weak assumptions concerning $h(.)$ and $f(.)$ the fractal embedding theorem [14]ensures that, for $D > 2d$, the set of all *delayed coordinate vectors*

$$Y_{D,T} = \{k > k_0 : \vec{y_k} = (y_k, y_{k-T}, \ldots, y_{k-(D-1)T})\}, \qquad (2)$$

with an arbitrary delay time T, forms an embedding of A in the D-dimensional *reconstruction* space. The minimal D, which yields an embedding of A, is called the *embedding dimension*. An embedding preserves the characteristic features of A, especially it is one to one, and therefore, can be employed for building a system model. For this purpose the reconstruction of the attractor is used to uniquely identify the systems state thereby establishing the possibility of uniquely predicting the systems evolution. By iterating this prediction function $f(\vec{y}_k) = y_{k+T}$ we obtain a system model which, however, is useful only in the neighborhood of the trained attractor. To obtain an analytic system model we use the attractor samples reconstructed from the time series to train a neural network as an estimator of the prediction function.

In the case of chaotic system dynamics this approach to modeling has been successfully investigated by a number of researchers [5; 1; 8]. For modeling musical instruments the argumentation takes over directly. The time series in this case is the sound signal, which is used to reconstruct the underlying attractor. Up to now, the neural network modeling of the sound dynamics has been successfully applied to the resynthesis of saxophone, piano and speech signals [10].

4 Morphing sound models

The new morphing algorithm we are going to present now, is based upon the homotopic mixing of dynamical systems [3]. In our approach we use as the progenitor systems the sound models obtained from the respective sounds. Then we construct a new sound model with an additional morphing parameter α that consists of the convex sum of the progenitor sound models, $f_1(.)$ and $f_2(.)$, following

$$f_m(\vec{y}, \alpha) = \alpha f_1(\vec{y}) + (1-\alpha) f_2(\vec{y}). \tag{3}$$

The morphing parameter α is confined to the interval $[0,1]$, such that the parameterized model $f_m(., \alpha)$ establishes a homotopic transformation between the progenitor models. For $\alpha = 1.0$ and $\alpha = 0.0$ the morphing model $f_m(., \alpha)$ reproduces the progenitor models, and for intermediate values of α new sound dynamics are produced. Smooth changes of α result in smooth changes of the dynamical model $f_m(.)$. However, smooth changes in $f_m(.)$ does not necessarily result in smooth changes of the attractor, because for varying α all kinds of bifurcations may occur [3]. To achieve reasonable interpolating dynamics the progenitor attractors have to be geometrically and topologically similar. While it is very difficult to quantify this similarity it seems necessary that the two progenitor attractors lie in the basin of attraction of each other. The following experiments are intended to shed some light on this topic and will demonstrate that the differences that yield smooth interpolations seems to be large enough to assume reasonable applications of the method.

Signal abbrev.	Fundamental $w_0/2/\pi$	Amplitude a_3	Phase ϕ_3	Frequency shift w_r	Attractor dimension
HAR	$\frac{100+\pi}{6000}$	0.3	0.0	1.0	1
PHA	$\frac{100+\pi}{6000}$	0.3	$\frac{\pi}{3}$	1.0	1
AMP	$\frac{100+\pi}{6000}$	0.7	0.0	1.0	1
HIG	$1.0595\frac{100+\pi}{6000}$	0.3	$\frac{\pi}{3}$	1.0	1
INH	$\frac{100+\pi}{6000}$	0.3	0.0	1.0595	2

Tab. 1: Parameter settings of eq. (4) for all synthetic signals that will be used in the following experiments.

5 Morphing synthetic sound signals

To obtain a fundamental understanding of the properties of the homotopic mixture of sound models we first investigate the results obtained for simple synthetic sound signals. All synthetic signals are comprised of 3 partials with varying phase, frequency and amplitude characteristics. The mathematical description of the signals is

$$y_n = \sin(w_0 k) + a_3 \sin(3 w_0 w_r k + \phi_3) + 0.2 \sin(5 w_0 k). \tag{4}$$

By varying the parameters w_0,a_3, ϕ_3 and w_r a number of interesting situations for the investigation of sound morphing may be established. All signals that will be used in the following experiments are listed in table (1). The fundamental frequency has been chosen to prevent any rational relation between signal and sample frequency f_a, because in this case the attractor would degenerate to be a collection of points.

The signals with $w_r = 1.0$ are periodic and, therefore, possess an attractor being a closed line. The signals *PHA* and *AMP* differ from *HAR* only by a change of the amplitude (a_3) or the phase (ϕ_3). To investigate the properties of the morphing algorithm for the interesting situation of progenitor signals with different pitch we employ signal *HIG*, which has a fundamental frequency w_0 one half tone above the other signals.

While all these signals have attractors with the same topology the last example *INH* is not periodic due to the shifted frequency of the third partial. Due to the non harmonic partial the topology of the attractor is changed from a closed line into a two-torus. Due to the change in topology, we consider the morphing between *INH* and the other signals the most difficult case. The attractor dimensions given in table (1) follow directly from the frequency characteristics of the signals [4].
For all the signals given in table (1) we trained dynamical sound models as explained in [10]. The neural networks used are of the radial basis function type with normalized hidden units. The network implements a function

$$\vec{N}(\vec{y}_k) = \sum_j \vec{w}_j \frac{\exp\left(-(\frac{\vec{c_j}-\vec{y}_k}{\sigma_j})^2\right)}{\sum_i \exp\left(-(\frac{\vec{c_i}-\vec{y}_k}{\sigma_i})^2\right)} + \vec{b}. \qquad (5)$$

The network parameters $\vec{w}$, $\vec{c}$ and $\vec{b}$ are adjusted by means of a standard training algorithm *RPROP* [7] to obtain optimal prediction of the following sample y_{k+T}. Iteration of the model and up sampling the output signal (to compensate for the step size T) yields a time series that, for a stable model, closely resembles the training signal. However, we note again that due to random effects during training and due to the fact that the training data is restricted to the attractor, a stable result is not guaranteed. Especially, for morphing applications where we drive the models away from its attractor the result may be undefined. For the present investigation with simple synthetic signals we trained 5 models for each signal. All models proved to be stable. For all signals except *INH* we choose the networks to consist of 10 input units, 20 hidden units and one output unit with a delay time $T = 4$. For modeling the signal *INH*, we have chosen the same network topology, however, due to the higher complexity of the attractor the number of hidden units has been increased to 60 in this case. For the simple signals used here recurrent training of the models has not been necessary.

In the following we will describe our results obtained for synthesizing time series from convex sums of the dynamical sound models just described. To be able to study the stationary characteristics of the mixed models we select the morphing parameter α to be fixed from the set $[0.0, 0.2, 0.4, 0.6, 0.8, 1.0]$.

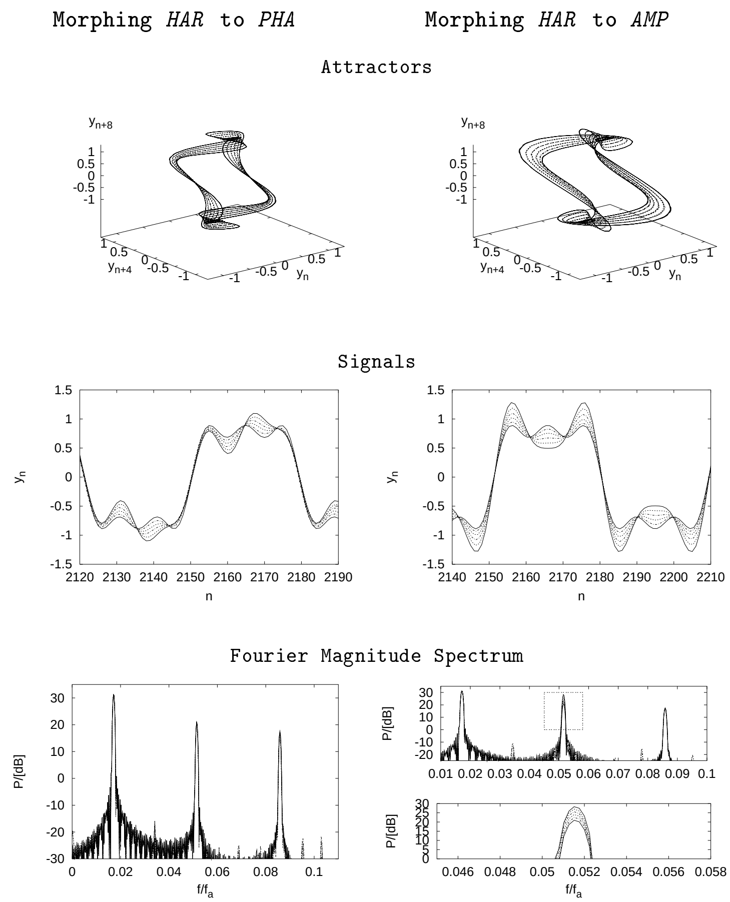

Fig. 1: Morphing results. The solid lines represent the cases $\alpha = 0.0$ and 1.0 while the dotted lines represent $\alpha = 0.2, 0.4, 0.6, 0.8$.

Both models are initialized by the first vector of one of the original time series and iterated following eq. (3). After the transients have decayed ($k = 3000$), we recorded further 6000 samples of the time series that are used for further analysis. In the left column of figure 1 we present the reconstructed attractors, short signal segments and Fourier spectra of the synthesized time series obtained from morphing between the models *HAR* and *PHA*. The attractor picture reveals how the parameter α drives the model $f_m(., \alpha)$ from one attractor to the other. This smooth geometrical transformation results in a smooth transformation of the time series. The magnitude of the Fourier spectrum, however, is nearly the same for all α. This is a nice result, because it is not an obvious property that the geometrical transformation obtained in the reconstructed state space yields Fourier spectra with constant spectral peaks. The nonlinear distortion of the morphed signals is very small,

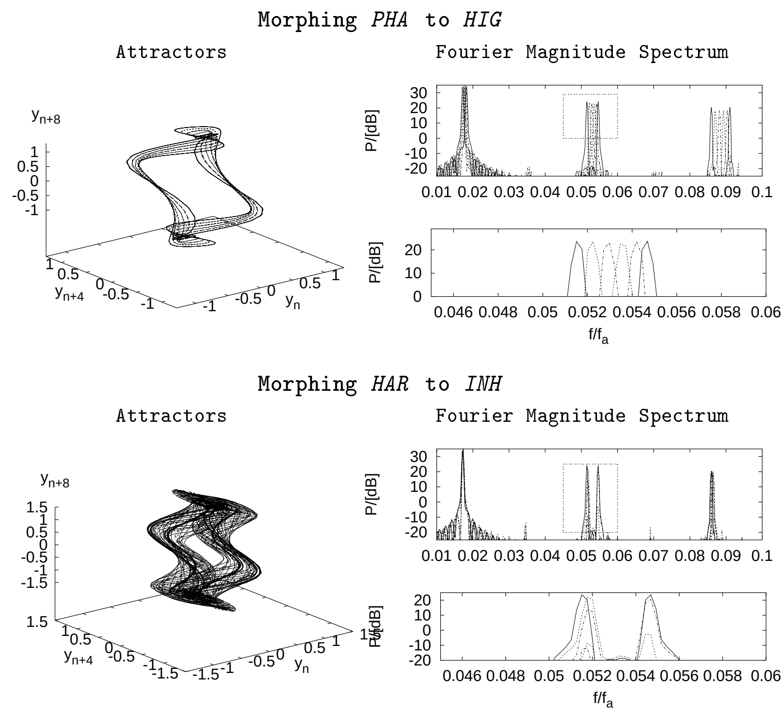

Fig. 2: Morphing results. Line types as in figure 1, values of α see text.

which can be concluded from the fact that any additional partials shown in the spectra have spectral power well below the partials of the progenitor signals. In the right column of figure 1 we present the same results for morphing between *HAR* and *AMP*. Again we find the satisfying result that the geometrical transformation yields Fourier spectra with stable spectral peaks, where only the magnitude of the third partial is increased during the morph.

In the top line of figure 2 we present the results for morphing between the signal *PHA* and *HIG* with α fixed to the same value as before. Due to the simple relation between time series and attractor we skip the presentation of time series here. The attractor figure looks quite similar to the results obtained from the morph between *HAR* and *PHA*. However, in this case the attractor belonging to *HIG* is passed through with higher velocity. Nevertheless, the geometrical transformation remains smooth. The Fourier magnitude spectrum reveals the fact that the geometrical morph shifts the frequencies of the individual partials such that the partials are always harmonic and the resulting signal is always periodic. Again the geometrical interpolation proves satisfactory with respect to the spectral characteristics of the morphed signals, because a glissando (smooth pitch change) of a natural instrument will generally exhibit the same spectral properties [15].

In the bottom row of figure 2 the results for the morph between *HAR* and

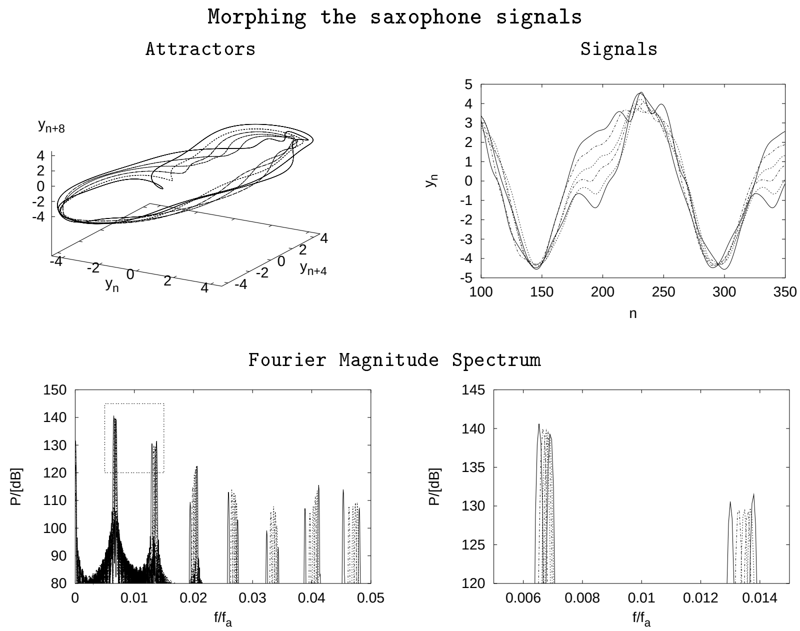

Fig. 3: Morphing results for saxophone signals. Line types as in figure 1.

INH are shown. The picture of the attractors shows the geometrical transformation, however, due to the two dimensional attractor of the signal *INH* the interpretation of the figure is rather difficult. Considering the spectrum we find that an increase of α from 0.0, where the signal is periodic, produces periodic solutions with slightly increased fundamental. Further increasing α reduces the magnitude of the third partial and, later, synchronously increases the magnitude of the non harmonic partial and decreases the fundamental again. In this last case the spectral properties obtained from the morph are not what one would expect from a morphing algorithm, because the main mechanism to achieve the morph is not a frequency shift, but, a complex process of frequency shifting and amplitude mixing. The situation of the last experiment, however, is to some extend not natural, because for a natural instrument it is generally not possible to shift the partials independently.

Note, that in the last example α has been set to $[0.0, 0.4, 0.8, 0.95, 0.98, 1.0]$, because the change of topology appears only for large values of α. A question that arise in this context is, whether the increase in dimension always needs a higher driving force, such that α has to be closer to the model with the higher dimensional attractor to obtain the higher dimension from the morphing model $f(.,\alpha)$.

6 Morphing of real world sounds

After having obtained encouraging result for morphing simple synthetic signals we additionally tested the algorithm using two saxophone signals that have been played by a professional player with similar excitation and pitch difference of one half tone. The signals with a fundamental frequency of about 330 Hz have been recorded with 48000 kHz sample frequency and 16 Bit resolution. The neural network models used in this experiment have 20 input units 120 hidden units and 1 output unit and the delay time is selected to be $T = 5$. For training the network we again use the *RPROP* algorithm, however, we employ 4 steps of recurrent backpropagation for each data sample to ensure stability of the trained models [5; 1].

The trained networks are able to resynthesize the training signals with high quality. The morphing algorithm is then applied to obtain morphed saxophone signals for the set of fixed α again equally space between zero and one. The results presented in figure 3 confirms the findings obtained for synthetic signals. Again the attractors of the saxophone signals are smoothly transformed, while their topology is preserved. Analyzing the Fourier spectrum reveals the striking fact that the partials of the intermediate signals are smoothly shifted, such that all intermediate signals preserve the characteristics of the original saxophone sound signals.

However, we have to note here that without additional constraints of the network parameters the morphed signals did exhibit a considerable amount of nonlinear distortion at high frequencies. This is due to the fact that some of the RBF units are located well apart from the attractor. If the model is operated in this regime the large distance to the other units leads to a prediction function that is locally constant, which in turn produce the undesirable distortion. The problem has been cured by formulating an additional constraint for the training algorithm, such that the width parameters of the RBF units σ_i are kept above a fixed value. With this constraint the distortions in the synthesized signals are no longer audible.

7 Summary

In the present article we have described a new algorithm for morphing natural sound signals by means of attractor morphing. We investigated some fundamental properties of the algorithm, that basically yields a geometrical interpolation of the progenitor sound attractors. The morphed signals obtained by this interpolation have natural spectral characteristics as long as the progenitor attractors are topologically equivalent. The findings are very promising in that the satisfactory spectral properties of the morphed sound signals are achieved by an algorithm that is strictly working in the time domain.

References

[1] G. Deco and B. Schürmann. Neural learning of chaotic dynamics. *Neural Processing Letters*, 2(2):23–26, 1995.

[2] A. Lapedes and R. Farber. How neural nets work. *IEEE Conference on Neural Information Systems*, 1:442–457, 1987.

[3] R. Mettin and G. Mayer-Kress. Chaotic attractors from homotopic mixing of vector fields. *International Journal of Bifurcation and Chaos*, 6(2):395–408, 1996.

[4] T. Parker and L. Chua. *Practical Numerical Algorithms for Chaotic Systems*. Springer-Verlag, New York-Heidelberg-Berlin, 1989.

[5] J. C. Principe and J.-M. Kuo. Dynamic modelling of chaotic time series with neural networks. In G. Tesauro, D. S. Touretzky, and T. Leen, editors, *Neural Information Processing Systems 7 (NIPS 94)*, 1995.

[6] J. C. Principe, A. Rathie, and J.-M. Kuo. Prediction of chaotic time series with neural networks and the issue of dynamic modeling. *Int. Jour. of Bifurcation and Chaos*, 2(4):989–996, 1992.

[7] M. Riedmiller. Advanced supervised learning in multilayer perceptrons – From backpropagation to adaptive learning algorithms. *Computer Standards and Interfaces, Special Issue on Neural Networks*, 5, 1994.

[8] A. Röbel. *Neural Models of Nonlinear Dynamical Systems and their Application to Musical Signals*. PhD thesis, Technical University of Berlin, 1993. In German.

[9] A. Röbel. Using neural models for analyzing time series of nonlinear dynamical systems. In *Proceedings of the 5th International IMACS-Symposium on System Analysis and Simulation*, 1995.

[10] A. Röbel. Neural network modeling of speech and music signals. In *Neural Information Processing Systems 9 (NIPS 96)*, 1997.

[11] D. Ruelle. Characteristic exponents for a viscous fluid subject to time dependent forces. *Commun. Math Phys.*, 93:285–300, 1984.

[12] D. Ruelle. Diagnosis of dynamical systems with fluctuating parameters. *Proc. of the Royal Society London*, 413:5–8, 1987.

[13] T. Sauer, J. A. Yorke, and M. Casdagli. Embedology. *Journal of Statistical Physics*, 65(3/4):579–616, 1991.

[14] F. Takens. *Detecting Strange Attractors in Turbulence*, volume 898 of *Lecture Notes in Mathematics (Dynamical Systems and Turbulence, Warwick 1980)*, pages 366–381. D.A. Rand and L.S. Young, Eds. Berlin: Springer, 1981.

[15] E. Tellman, L. Haken, and B. Holloway. Timbre morphing of sounds with unequal numbers of features. *Journal of the Audio Engineering Society*, 43(9):678–689, 1995.