



HAL
open science

Morphing Sound Dynamics

Axel Roebel

► **To cite this version:**

Axel Roebel. Morphing Sound Dynamics. 1998 International Conference on Neural Networks and Brain, 1998, Beijing, China. <hal-02911713>

HAL Id: hal-02911713

<https://hal.science/hal-02911713v1>

Submitted on 4 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Morphing Sound Dynamics

Axel Röbel

Technical University Berlin, Sekr. EN-8, Einsteinufer 17,
10587 Berlin, Germany

Tel: +49-30-314 25699, FAX: +49-30-314 21143,

email: roebel@kgw.tu-berlin.de

Abstract

We investigate a new algorithm for sound morphing that is strictly working in the time domain and is based on a recent method of modeling natural sounds with neural networks called dynamic modeling. By means of basic synthetic signals we investigate the fundamental properties of the proposed morphing scheme. A successful morph of real world sounds of a musical instrument demonstrates the practical application. The results are particularly satisfactory because the time domain algorithm show frequency domain characteristics that are desirable for natural sound morphing applications.

1 Introduction

The use of neural networks for the modeling of time series is one of the important applications of the theory of nonlinear dynamical systems in the field of neural networks. The basic mathematical foundation for this application, the reconstruction theorem, has been formulated first in 1980 [13, 12]. Recently it has been discovered that the 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 [5, 8]. This application is generally termed dynamic modeling. 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 [9]. 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, we follow another approach to achieve signal modification, which we will call attractor morphing. This approach is motivated by a recent investigation on homotopic mixture of vector fields [2], which demonstrates the generation of new system dynamics by means of a

homotopic mixture of chaotic progenitor systems. In our investigation we have shown 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 signals obtained by means of attractor morphing achieve the characteristics of natural sounds without using frequency domain information [14]. We conclude that due to the close relations to the physical dynamics the attractor modeling of sounds has favorable properties for modeling natural sounds.

2 Algorithm

Traditional musical instruments belong to the class of dissipative, nonlinear mechanical systems. The discrete time evolution of such a system may be described in a k -dimensional state space S by means of a mapping $f(\cdot)$

$$\vec{z}_{n+1} = f(\vec{z}_n) \quad \vec{z} \in S \quad , \quad (1)$$

which connects the system state \vec{z} at time n , with the system state at the next time step. For a stable system $f(\cdot)$ and for large $n \rightarrow \infty$ the state \vec{z}_n will be confined to a bounded and closed subset $A \subset S$ of the state space, which is called an *attractor* of $f(\cdot)$ [10]. An attractor may be as simple as a point or as complex as a fractal set if the system dynamics are chaotic. If the dynamical system generates an output (sound) signal y_n , the characteristics of this signal are closely related to the topology of the attractor [3].

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. (5) for all synthetic signals that will be used in the following experiments.

For a d -dimensional attractor generating a sound signal y_n the fractal embedding theorem [13, 12] ensures under weak assumptions concerning y_n and $f(\cdot)$, that for $D > 2d$, the set of all *delayed coordinate vectors*

$$Y_{D,T} = \{n > n_0 : \vec{y}_n = (y_n, y_{n-T}, \dots, y_{n-(D-1)T})\}, \quad (2)$$

with an arbitrary delay time T , forms an embedding of A in the D -dimensional *reconstruction* space G . Because an embedding preserves the characteristic features of A , especially it is one to one, it may be employed for building a system model. To achieve this we use a neural network to approximate the system function $f(\cdot)$ in the reconstructed state space G , yielding a prediction model

$$y_{n+T} = f_N(\vec{y}_n). \quad (3)$$

For stable predictor models f_N the prediction can be iterated thereby establishing a model of the system dynamics on attractor A . The building of predictor based neural network dynamical models has been successfully applied to chaotic system dynamics [4, 1] and also to non chaotic sound dynamics of saxophone, piano and speech signals [9]. Note, however, that in the case of music and speech signals the system dynamics are generally not stationary. For slowly varying dynamics this situation can be described by a system undergoing a parameter variation and, therefore, following a sequence of attractors [11, 7].

3 Morphing sound models

The new morphing algorithm we are going to present now, is based upon the homotopic mixing of dynamical systems [2]. We use as progenitor systems the sound models obtained from the respective sounds and construct a new sound model with an additional morphing parameter α that consists of the convex sum of the progenitor sound models, $f_1(\cdot)$ and $f_2(\cdot)$, following

$$f_m(\vec{y}, \alpha) = \alpha f_1(\vec{y}) + (1 - \alpha)f_2(\vec{y}). \quad (4)$$

The morphing parameter α is confined to the interval $[0, 1]$, such that the parameterized model $f_m(\cdot, \alpha)$ establishes a linear transformation between the progenitor models. For $\alpha = 1.0$ and $\alpha = 0.0$ the morphing model $f_m(\cdot, \alpha)$ reproduces one of the progenitor models, and for intermediate values of α new sound dynamics are produced. Smooth changes of α results in smooth changes of the dynamical model $f_m(\cdot)$. However, smooth changes in $f_m(\cdot)$ does not necessarily result in smooth changes of the attractor, because for varying α all kinds of bifurcations may occur [2]. To achieve interpolating dynamics with the desired acoustical properties the progenitor attractors have to be sufficiently similar. While it is very difficult to quantify this similarity it seems necessary that the progenitor attractors lie in the basin of attraction of each other.

Note that, due to the fact that the training data is restricted to the related attractor, a sensible model behavior is generally restricted to a close neighborhood of the attractor. This might become a problem for a morphing application if the morph drives the model to far from its attractor.

4 Results

To obtain a fundamental understanding of the properties of the homotopic mixture of sound models we tried the algorithm for morphing between a number of artificial sound signals given by

$$y_n = \sin(w_0 n) + a_3 \sin(3w_0 w_r n + \phi_3) + 0.2 \sin(5w_0 n). \quad (5)$$

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.

For all the signals given in table (1) we trained dynamical sound models as explained in [9]. The neural networks used are of the radial basis function type with normalized hidden units. We select the number of hidden units to be high enough (60 for *INH*

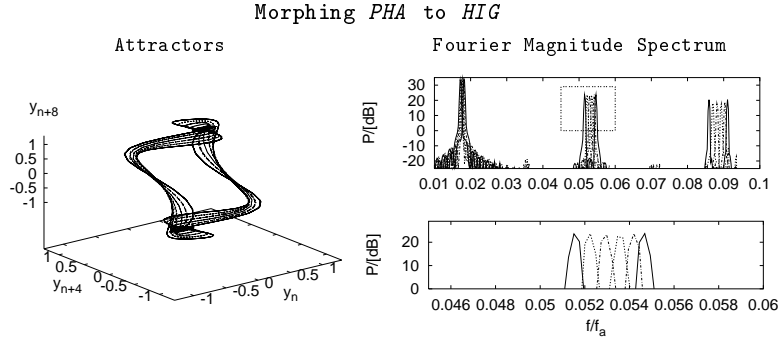


Fig. 1: Morphing signal *PHA* into *HIG*. The solid lines represent the cases $\alpha = 0.0$ and 1.0 while the dotted lines represent $\alpha = 0.2, 0.4, 0.6, 0.8$.

and 20 in all other cases) to achieve stable iteration of the model. Because the delay time T determines the down sampling of the time series that is generated by the iterated model it should be chosen to lie below the Nyquist sampling period. Therefore, we choose $T = 4$. To prevent local loops in the reconstructed attractor we used input vectors \vec{y}_n that span at least half a period of the time series. Therefore, we have chosen the number of input units to be 10 which is well above the required embedding dimension D .

As an example we consider here the case of morphing between signal *PHA* and *HIG*. After having trained dynamical models for both signals we used eq. (4) with fixed α , $0.0 \leq \alpha \leq 1.0$, to synthesize sound signals with interpolating dynamics. The evolution of the morphing model in a reconstructed state space shown in figure 1 reveals how the parameter α drives the model $f_m(\cdot, \alpha)$ from one progenitor attractor to the other. This smooth geometrical transformation results in a smooth transformation of the time series. To assess the sound characteristics of the morphed signals we analyze their Fourier spectra. As shown in figure 1 we find 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. This is the kind of behavior one would generally expect from a natural glissando.

These results are particularly satisfactory because the morph with glissando like spectral properties of the morphed signals have been obtained by means of an algorithm that is strictly working in the time domain. We have obtained this type of result in all cases where the progenitor sound attractors are topological equivalent, that is they have the same local dimension. In case of a morph between the artificial signals *PHA* and *INH*, which have topologically different attractors, the morphing results did not show the simple natural characteristics. This setting, however, is itself not natural because for a musical instrument it is gen-

erally impossible to shift a single partial.

4.1 Application to saxophone sound signals

To verify our results we used two segments of real world 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 morphing algorithm is applied to obtain morphed saxophone signals with different pitch for the set of fixed α . The results presented in figure 2 confirms the findings obtained for synthetic signals. Again the attractors of the saxophone signals are smoothly transformed, while their topology is preserved. The Fourier spectra reveal the 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 saxophone signals did exhibit a considerable amount of nonlinear distortion at high frequencies. This is due to the fact that some of the RBF units of each model are located well apart from the attractor. If the model is operated in this regime the large distance to the other units and the normalization of the hidden layer 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 are kept above a fixed value. With this constraint the distortions in the synthesized signals are no longer audible.

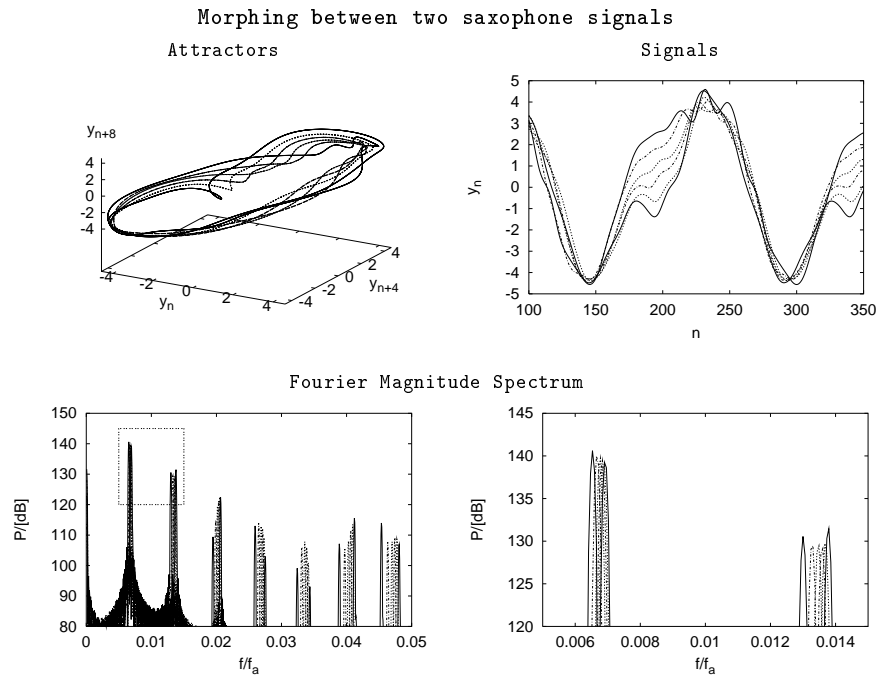


Fig. 2: Morphing results for saxophone signals. Line types as in figure 1.

References

- [1] G. Deco and B. Schürmann. Neural learning of chaotic dynamics. *Neural Processing Letters*, 2(2):23–26, 1995.
- [2] 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.
- [3] T. Parker and L. Chua. *Practical Numerical Algorithms for Chaotic Systems*. Springer-Verlag, New York-Heidelberg-Berlin, 1989.
- [4] 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.
- [5] 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.
- [6] A. Röbel. Morphing dynamical sound models. In *Proceedings of the 1998 IEEE Workshop on Neural Networks for Signal Processing VIII*, 1998. to appear.
- [7] A. Röbel. *Neural Models of Nonlinear Dynamical Systems and their Application to Musical Signals*. PhD thesis, Technical University of Berlin, 1993. In German.
- [8] 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.
- [9] A. Röbel. Neural network modeling of speech and music signals. In *Neural Information Processing Systems 9 (NIPS 96)*, 1997.
- [10] D. Ruelle. Characteristic exponents for a viscous fluid subject to time dependent forces. *Commun. Math Phys.*, 93:285–300, 1984.
- [11] D. Ruelle. Diagnosis of dynamical systems with fluctuating parameters. *Proc. of the Royal Society London*, 413:5–8, 1987.
- [12] T. Sauer, J. A. Yorke, and M. Casdagli. Embedology. *Journal of Statistical Physics*, 65(3/4):579–616, 1991.
- [13] 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.
- [14] 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.