# Multi-Touch Enhanced Visual Audio-Morphing

Lars Engeln, Dietrich Kammer, Leon Brandt, Rainer Groh
Technische Universität Dresden
01062 Dresden, Germany
firstname.lastname@tu-dresden.de

## ABSTRACT

Many digital interfaces for audio effects still resemble racks and cases of their hardware counterparts. For instance, DSP-algorithms are often adjusted via direct value input, sliders, or knobs. While recent research has started to experiment with the capabilities offered by modern interfaces, there are no examples for productive applications such as audio-morphing. Audio-morphing as a special field of DSP has a high complexity for the morph itself and for the parametrization of the transition between two sources. We propose a multi-touch enhanced interface for visual audio-morphing. This interface visualizes the internal processing and allows direct manipulation of the morphing parameters in the visualization. Using multi-touch gestures to manipulate audio-morphing in a visual way, sound design and music production becomes more unrestricted and creative.

## Author Keywords

audio, visual audio, morphing, gesture interaction, touch

## CCS Concepts

•**Applied computing** → **Sound and music computing;** •**Human-centered computing** → *Gestural input;* User interface design;

## 1. INTRODUCTION

Audio-morphing by itself is a vague term [5]. Creating hybrid sounds from two sources as well as transitioning a sound's timbre from one to another are both often called *audio-morphing*. For a clear distinction, we use the term *morphing* to refer to a transition of the timbre over time. The result of this process is called *morph*. *Hybrids* instead are new sounds made from existing ones.

Audio-morphing is achieved using different approaches [17, 3, 9, 13]. The most common way is to map partials. However, this sequential mapping does not always yield the most fluent morph. Partials are therefore mapped to the nearest neighbour [16], to minimize the partial movements across the frequencies. Remaining partials are interpolated to phantom-partials with zero amplitude (so called *Zero-amplitude partials* [16]), which are harmonics of the fundamental tone.


*NIME'18,* June 3-6, 2018, Blacksburg, Virginia, USA.

Partials for hybrids and morphs are analyzed and manipulated within the frequency-domain as well as resynthesized afterwards. In addition, there are tools and papers concerning generally visual manipulation of the frequency-domain [4, 11] and the timbre design [6, 15]. Our research describes a more comprehensible workflow for sound *designers* to creatively manipulate and generate sounds.

In case of visual audio-morphing, interesting related work has been described by Efros and Leung concerning texture synthesis and completion [7]. Barnes et al. address *guided* hole-filling algorithms for images [1].

In contrast, our work focuses on visually manipulating partial mapping and interpolation. To this end, we developed a gesture set for touch input. Note that keeping track of background-noise or structural morphing of images is out of the scope of this paper.

## 2. VISUAL AUDIO-MORPHING

To morph a source $A$ into a source $B$, three steps are required: detection, mapping, and interpolation of the partials. Subsequently, a (re-)synthesization transforms back to the time-domain. For audio-morphing, various parameters should be regarded. Our gesture set described in section 2.2 is designed to interact with these parameters. The use of multi-touch gestures led to novel interaction possibilities to shape a morph that we discuss in section 2.3.

### 2.1 Parameters

The interpolation parameters can be divided in the following categories.

- General temporal aspects of the morph (section 2.1.1)
- Partial characteristics for interpolation (section 2.1.2)
- Registration of partials to each other (section 2.1.3)

The 10 parameters are shown in Figure 1 and referenced in the following subsections via numbers in parentheses.

#### 2.1.1 Temporal Aspects

The length of a morph is defined by the end of $A$ and start of $B$ (see (1) in Figure 1). The amount of data (time of a source), which is considered for the interpolation (2) is important to preserve features such as tremolo and other variations. At the same time, it suppresses possibly unwanted features such as attack and decay.

#### 2.1.2 Partial Characteristics

Parameters concerning a partial besides magnitude and frequency (3) are needed to perform a smooth transition from a partial of $A$ to a partial of $B$. So far, all variations over frequency and time like magnitude alterations (4) and tremolo (5) belong to the basic parameters. Moreover, a partial's

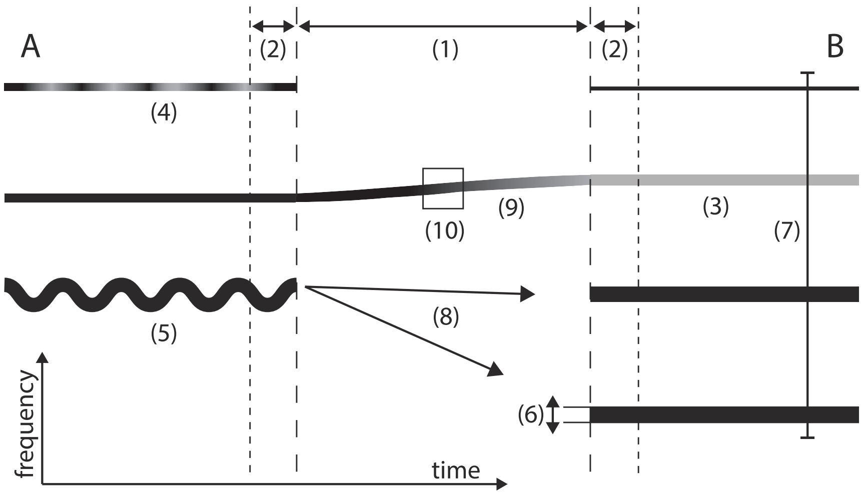


**Figure 1: Basic parameters for audio-morphing. (1) transition time from A to B, (2) considered data to morph, (3) magnitude, (4) magnitude alterations, (5) tremolo, (6) (band-)width and band border smoothness, (7) amount of partials, (8) mapping of partials, (9) interpolation speed, (10) turning point of interpolation curve**

bandwidth and band border behavior (e.g. smoothness) (6) are discernible.

### *2.1.3 Registration*

To register two sources $A$ and $B$, the partials of both are mapped to each other (8). The target for a smooth morph is a partial ratio of $1:1$ and for more creative, likewise noisy morphing, it can have arbitrary partial ratios $(n:m)$. Partials with no mapping are interpolated to *Zero-amplitude partials* [16], which are harmonics of the fundamental. The mapping itself can be sequential, the nearest neighbour [16], or any other function (e.g. $A(x) \rightarrow B(2x)$ for skipping odd partials of B). The interpolation along a spline-curve has a turning point (10). The position and the tangent of the turning point shapes the way of interpolation, whether it is a steady or fast transition from $A$ to $B$ for instance.

## 2.2 Gesture Set

Wobbrock et al. conducted a user elicitation study for tabletop touch gestures [18, 12]. In the study, different tasks were presented that should be solved using gestures. Thereby, the most intuitive gestures were elicited from the study participants. Afterwards, Wobbrock et al. composed a conflict-free gesture set from the data. In this section, we use the terminology and basic gesture set introduced by these authors. Although the gesture set is verbalized for more readability, it could be directly transformed into *GeForMT* [10] syntax, a modern gesture formalization for multi-touch. Our gesture set shown in Figure 2 is context-sensitive and distinguishes between object manipulations and workspace gestures. Local object gestures manipulate the object itself (e.g. the source sound A) and global gestures affect the entire workspace. Global gestures are one-hand-move (at least 3 fingers moving on the surface) for panning the workspace (see (2) in Figure 2) and one-hand-splay (at least 3 finger performing a pinch) for zooming the workspace. Moreover, panning with one finger and pinching with two fingers performed on an empty area of the workspace (1) pans and zooms the workspace. In the following, the object gestures for the parameters (see section 2.1) are discussed. Then, gestures for multiple turning points are outlined.

### *2.2.1 Temporal Aspects*

The position in time of the sources A and B are modified with a one-finger-move (3), thereby increasing the duration

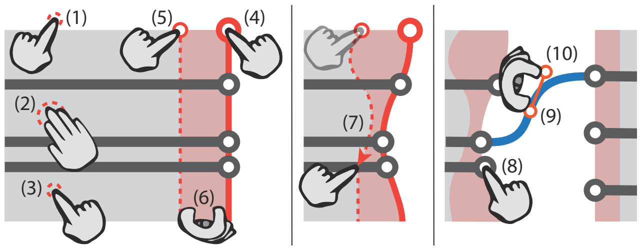


**Figure 2: Basic gestures. (1) workspace pan (empty area), (2) workspace pan (everywhere), (3) position of the sound, (4) trims the sound, (5) modifies the considered data, (6) enlarges the considered data-range with a pinch, (7) freehand manipulation of (5) and (6) (on specific borderline), (8) selects partial, (9) one-finger position of turning point, (10) two-finger position of turning point and tangent's slope).**

of the morph. If the one-finger-move is performed at the end of A or at the start of B (4), the sources are trimmed. To select the considered data for the morphing (cf. section 2.1.1), a one-finger-move is performed on the top of the specific border (5). In addition, a pinch enlarges the range of the considered data (6). Furthermore, a one-finger-move shapes the borders of the trimming and the considered data freely, if the gesture does not start at the very top (7).

### *2.2.2 Partial Characteristics*

Since all characteristics of the partials (see section 2.1.2) are used for the interpolation of the transition, no gestures are applied on these directly. The horizontal position of the turning point determines when 50% of the interpolation is done. Via a one-finger-move, the position of the turning point is changed (9). A second finger modifies the slope of the tangent (one-finger-hold + one-finger-move, comparable with a two-finger-pinch) (10). By adjusting the tangent, the way of interpolation is shaped significantly.

### *2.2.3 Registration*

At first, a partial is selected with a tap at the partial's end point (8). Multiple partials are selected either with a lasso gesture, with a one-finger-hold at the first one and tapping at the others, or with a one-finger-hold at the first and a

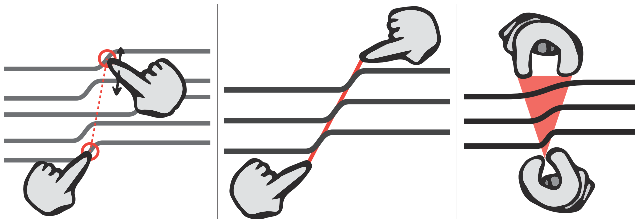

**Figure 3: Gestures for fast manipulations of multiple turning points. Selection of multiple turning points and alignment between the touch points (left), increasing the temporal intervals of the aligned turning points (middle), asynchronous change of the transition speed (slope of the tangents) (right).**

double-tap at the last to select all partials in between. Afterwards, a drag and drop registers the selected partials to the corresponding partial of the other source. Alternatively, a flick gesture automatically maps the selected partials (or all, if none was selected) to the nearest-neighbour. If an initial registration exists, the mapping follows that pattern. For instance, to skip the odd partials $P$ of source B: manually add $P_A(0) \rightarrow P_B(0), P_A(1) \rightarrow P_B(2)$, then flick to automatically map $P_A(2)$ to $P_B(4)$ and so on.

Moreover, a double-tap on an interpolation line deletes this particular registration and a double-tap on a partial deletes all of its corresponding registrations.

### 2.2.4 *Gestures for Multiple Turning Points*

Multi-touch enhanced visual audio-morphing can increase the speed of work significantly, because multiple turning points can be manipulated at once. In the following, we describe suitable gestures for that purpose.

The selection of multiple turning points is the same as it is for the partial ending points (see section 2.2.3, same meaning, same gesture). With a simultaneous long-tap at the first and a long-tap at the last turning point, all turning points are aligned between the touch points. Also, the alignment can be combined with the selection. Through a one-finger-hold and a double-tap-hold all turning points are selected and aligned at once (Figure 3, left). Furthermore, the tangents are blended as well. For instance, given are three tangents with slopes of 0°, 8°, and 30°. After the alignment, the second tangent in the middle is set to 15°.

While moving the fingers, the temporal intervals of the turning points are increased or decreased (Figure 3, middle). Thereby, the interpolation center of the individual transitions are either synchronous at one time, or asynchronous ascending or descending respectively. Moreover, the positions and the tangents are manipulated at the same time due to a bimanual gesture. Each hand uses two fingers to define a polygon. The polygon can vary from a rectangle, to a triangle, via a trapeze. The vertical positions align to the polygon's horizontal middle line and the tangents to the polygon's borders (Figure 3, right).

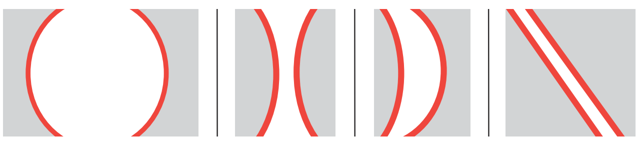

**Figure 4: Temporal Flavours. Examples of different border shape combinations.**

## 2.3 Different Flavours of Morph

Based on our work towards directly manipulating morphs, we identified novel possibilities to shape morphs in a visual way. We call the different possibilities *flavours*. Different flavours occur in particular through the mapping (see sections 2.1.3 and 2.2.3) and the interactions, which manipulate the turning points (see section 2.2.4). Every combination of the mappings, the positions of the turning points, and tangent shapes are possible. The freely manipulated borders of the temporal aspects (see section 2.2.1) have certain flavours as shown Figure 4. Different frequency bands can have different morph times. For instance, low and high partials remain original, while partials in the mid are already interpolated (see Figure 4, left). The mapping (e.g. 1 : 1 (each partial is mapped to only one other), nearest neighbour, arbitrary $n : m$ (each partial is mapped to none or more)) goes best with nearest neighbour, since beats are less likely to occur. If multiple partial lines diverge or merge in case of arbitrary mappings, noticeable beats are generated. However, a deliberate use of those beats can have an impressive effect. For example, if $P_A(0) \rightarrow P_B(m-1), P_A(1) \rightarrow P_B(m-2)$, ..., $P_A(n-1) \rightarrow P_B(0)$, then all partials melt into a center frequency and emerge with a band limited noise again.

The partial's interpolation center (turning point) can be consecutively ascending, synchronous at the same time, or consecutively descending (cf. Figure 5, upper left rows), besides other noisier or random arrangements. The tangent size affects the morph's speed (cf. Figure 5, upper left columns). Furthermore, the different interpolations can have arbitrary morph speeds within the frequencies (cf. Figure 5, bottom left, see bimanual gesture in section 2.2.3). Not all interpolations have to be aligned. Especially the interpolation of the fundamental as the base frequency can be treated differently than the other partials (cf. Figure 5, right). The fundamental's transition is displaced temporally and differed in speed. In this way, the timbre is morphed earlier or later, and faster, or longer. For example, the morphing of the timbre is already done, while the fundamental remains the same (cf. *temporal flavours*).

## 2.4 Proof-of-Concept

A 27 inch touch monitor has been used to test the basic gestures (see Figure 6). The implementation is written in C++ and is based on the libCinder framework [2]. If phantom-partials (see section 1) are switched on, partials with no mapping are interpolated to an even harmonic of the fundamental frequency. To transform back and forth between

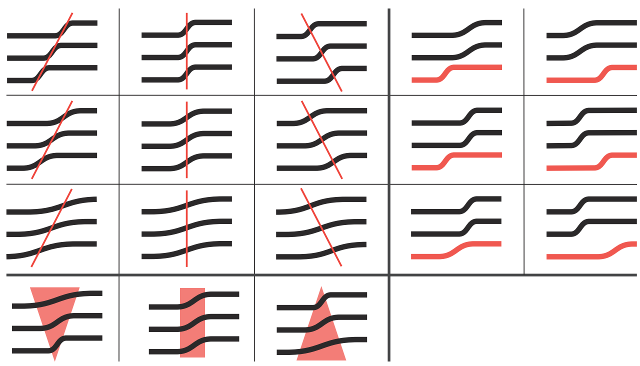

**Figure 5: Flavours of morph. Ascending and descending with slow and fast morphs (upper left), different morph speeds within the frequencies (bottom left), displaced fundamental (right).**

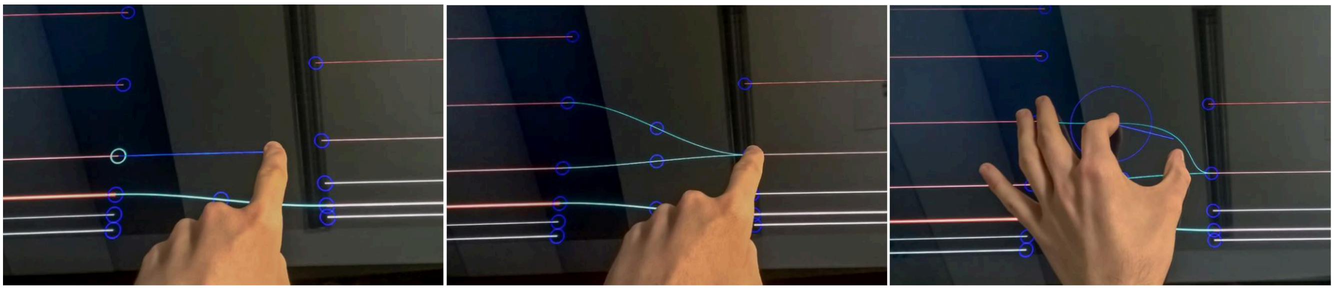

**Figure 6: Implementation: Gestures for applying manual mapping (left) and adjusting spline-curve (right)**

the time-domain and the frequency-domain the *Constant Q-Transform* [14] (CQT) is used. The CQT has a logarithmic spaced output for the frequency-domain in contrast to the linear spaced frequencies of the pure FFT. Thereby, the CQT's output fits the perception of the frequencies more accurately. Besides, for the equivalent results the CQT needs smaller computation windows than the FFT with binning (cf. *uncertainty principle* [8]).

## 3. CONCLUSIONS AND FUTURE WORK

Audio design using visual techniques is a long known research field. However, highly evolved hardware as well as algorithms are hardly addressed. Especially for effects such as morphing with an high amount of parameters a visual manipulation can be more efficient and comprehensible.

With the proposed gesture set based on Wobbrock et al. [18] findings, an intuitive way for creative morphings is achieved. Although an evaluation via user study is not conducted yet for this special set. A first attempt with subjects indicates a pleasurable handling of morphings in a playful manner. Besides an evaluation of the whole gesture set, the use-case of live performances should be considered as well. At the moment the workflow only addresses studio environments.

Performing image morphing [19] as an attempt for *real* visual audio-morphing is a next step. Therefore algorithms for structural image editing like *PatchMatch* [1] will be considered, because it suits the structure of the partials. In future work, we will head toward novel effects and interactions within the frequency-domain for productive and creative work.

## 4. REFERENCES


[1] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. 28(3):21–24, 2009.

[2] A. Bell, H. Nguyen, R. Eakin, and P. Houx. libCinder https://libcinder.org/, retrieved on 2018/01/20.

[3] F. Boccardi and C. Drioli. Sound morphing with Gaussian mixture models. In *Proc. DAFx*, pages 44–48, 2001.

[4] N. Bogaards, A. Roebel, and X. Rodet. Sound Analysis and Processing with AudioSculpt 2 To cite this version : Sound Analysis and Processing with AudioSculpt 2. *International Computer Music Conference (ICMC)*, (Icmc):1, nov 2004.

[5] M. F. Caetano and X. Rodet. Automatic timbral morphing of musical instrument sounds by high-level descriptors. In *International Computer Music Conference*, pages 11–21, 2010.

[6] G. Eckel. Manipulation of Sound Signals Based on Graphical Representation-A Musical Point of View. In *Proceedings of the International Workshop on Models and Representations of Musical Signals, Capri, Italia*, 1992.

[7] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1033–1038. IEEE, 1999.

[8] D. Gabor. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.

[9] T. Hikichi and N. Osaka. *Sound timbre interpolation based on physical modeling.* Acoustical Society of Japan, 2001.

[10] D. Kammer, D. Henkens, C. Henzen, and R. Groh. Gesture formalization for multitouch. *Software: Practice and Experience*, 45(4):527–548, 2015.

[11] M. Klingbeil. Software for spectral Analysis, Editing, and synthesis. In *ICMC*, 2005.

[12] M. R. Morris, J. O. Wobbrock, and A. D. Wilson. Understanding users' preferences for surface gestures. *Proceedings of Graphics Interface 2010*, pages 261–268, 2010.

[13] C. Pendharkar, M. Gurevich, and L. Wyse. Parameterized morphing as a mapping technique for sound synthesis. In *Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06)*, 2006.

[14] C. Schörkhuber and A. Klapuri. Constant-Q transform toolbox for music processing. *7th Sound and Music Computing Conference*, (january):3–64, 2010.

[15] A. Seago. A new interaction strategy for musical timbre design. In *Music and human-computer interaction*, pages 153–169. Springer, 2013.

[16] W. A. Sethares, A. J. Milne, S. Tiedje, A. Prechtl, and J. Plamondon. *Spectral Tools for Dynamic Tonality and Audio Morphing.* MIT Press - Journals, 2009.

[17] M. Slaney, M. Covell, and B. Lassiter. Automatic audio morphing. In *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, volume 2, pages 1001—-1004 vol. 2. IEEE, 1996.

[18] J. O. Wobbrock, M. R. Morris, and A. D. Wilson. User-defined gestures for surface computing. *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, page 1083, 2009.

[19] G. Wolberg. Image morphing: a survey. *The visual computer*, 14(8):360–372, 1998.