A FORMAL EVALUATION FRAMEWORK FOR SOUND MORPHING

Marcelo Caetano
FORTH-ICS
Heraklion, Crete, Greece
caetano@ics.forth.gr

Naotoshi Osaka
Tokyo Denki University
Tokyo, Japan
osaka@im.dendai.ac.jp

ABSTRACT

Sound morphing figures prominently as one of the most interesting sound transformation techniques due to its enormous creative potential. Most authors pose the problem of morphing sounds using perceptual requirements, but hardly ever evaluate their results mainly because perceptual evaluations are cumbersome and costly, and there are no standard objective evaluation criteria established for sound morphing. In this work we propose a formal evaluation procedure for sound morphing algorithms following three criteria, namely correspondence, intermediateness, and smoothness. The adoption of the proposed evaluation framework will help formalize the results towards more perceptually relevant morphed sounds.

1. INTRODUCTION

Sound morphing encompasses several models and techniques whose common goal is to obtain gradual transformations between sounds. Many sound transformations are called morphing in the literature [3], ranging from music compositions [10, 8] to the design of sound synthesizers [14] and the study of timbre spaces [2, 7]. It is very challenging to control the morph with a single parameter α, called morphing or interpolation factor [3, 4]. Ideally, the morphed sound should be perceived as halfway between source and target when $\alpha = 0.5$, for example.

Most morphing techniques proposed in the literature [1, 5, 6, 11, 12, 13, 14] use the interpolation principle, which consists in interpolating the parameters of the model used to represent the sounds and underplays the perceptual impact of the result. However, most authors pose the problem of morphing sounds using perceptual requirements, but hardly ever perform perceptual evaluations of their results mainly because perceptual evaluations are cumbersome and costly, and there are no standard evaluation criteria established for sound morphing. Usually, works about morphing sounds [1, 5, 6, 13, 14] present sound examples (commonly spectrograms) as results. Very seldom do the authors of these works present any evaluation. In this work we argue that sound morphing lacks a formal theoretical approach to properly evaluate the results. Following Osaka [12], we propose the following three criteria, correspondence, intermediateness, and smoothness. The next section briefly reviews sound morphing. Next, we discuss the evaluation criteria proposed, followed by subjective and objective evaluation procedures. Finally, we present the conclusions and future perspectives.

2. MORPHING SOUNDS

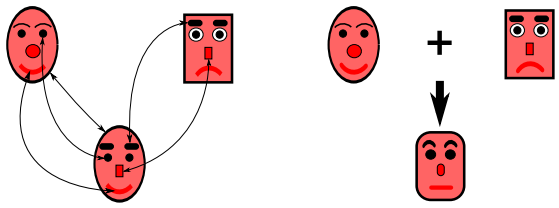
The aim of sound morphing is to obtain results that are perceptually intermediate between two (or more) sounds. The classic morphing technique is based on the interpolation principle, which suposes that we should obtain a gradual transition between the sounds if we interpolate the parameters of their representation. This assumes that there exists a sequence of intermediate sounds that will be perceived as a gradual transition between source and target. However, the perceptual impact of the interpolation of parameters depends largely on what information the parameters represent and how the sound material is perceived. On the one hand, if the parameters encode perceptually irrelevant information, the result of the interpolation of these parameters will very likely have little perceptual significance. On the other hand, if sound perception is categorical, we would try to achieve an impossible perceptually continuous transformation.

Morphing can be viewed as a transformation that involves hybridization to obtain intermediate form (among other features) [12]. The term hybridization is applied in many areas generally to refer to a process that involves the combination of two (or more) objects, individuals, varieties, etc, depending on what is being considered. A first important aspect of the problem of hybridization is to understand that there are several possible ways of combining two things. We are going to consider two simplified hybridization processes, one commonly found in nature and the other one usually only accessible by artificial means, which we call morphing.

2.1. Hybridization and Morphing

The hybridization process called sexual reproduction is ubiquitous in nature. In general terms, when a couple has children, we can usually easily recognize who the parents are because of physiological similarities. In other words, the kids take after their parents, and we usually say that they might have their mother's nose, the father's eyes, etc. This is a specific case of hybridization where the hybrid individual (the child) consists of a combination of parts from either parent, illustrated in figure 1a .

On the right, figure 1b depicts the process called



(a) hybridization (b) morphing

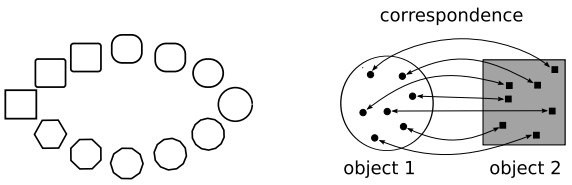
Figure 1: Depiction of face hybridization and face morphing. Part (a) shows the hybridization case where each attribute is inherited either from the mother or from the father. Part (b) shows a face whose constituent parts (nose, mouth, etc) are a combination of the corresponding parts from both parents and considered intermediate.

morphing, where each constituent part is now a combination of the corresponding parts from the parents. Therefore, each part is intermediate in shape (among other features). Figure 2a illustrates that sometimes there is more than one possible way of defining intermediate (or morphed) shapes. What is the intermediate shape between the circle and the square? When we are only considering shape, both transformations shown in figure 2a intuitively satisfy the requirement of a gradual transformation. The "best" or "most appropriate" transformation is application-dependent when we use objective criteria to evaluate the transformation or user-dependent when we use subjective criteria, that is, a user's personal taste or aesthetics.

2.2. Subjectiveness

Conceptually, the main difficulty in morphing is probably the fact that we are usually looking for a result that only exists as an abstraction. In other words, morphing allows to obtain the previously intangible, only accessible to our imagination. As we go up the complexity of forms and shapes, adding texture, colors, and other attributes, we are faced with questions such as: "What is the result of the morph between a tiger and a car?"

Sound perception is complex and abstract, rendering sound morphing very subjective. For example, we could try to imagine what the result of a morph between a dog bark and a trumpet note would sound like, but it is difficult to evaluate the results. What are the qualities that we expect to find in a good morph? Listeners are likely to be disappointed and give a low score when assessing morphed sounds simply because they do not meet their expectations, independent of the quality of the transformation. This raises the question of how to objectively evaluate the morphed sounds [12]. This work proposes three independent criteria: correspondence, intermediateness, and smoothness. In the next section, we will use visual analogies to illustrate each of them.



(a) shape transformation (b) correspondence

Figure 2: Depiction of shape transformation and correspondence. Part (a) illustrates two different possible paths that contain shapes that could be considered intermediate between the square and the circle. Part (b) illustrates that the morph depends on correspondence between elements.

3. EVALUATION CRITERIA

3.1. Correspondence

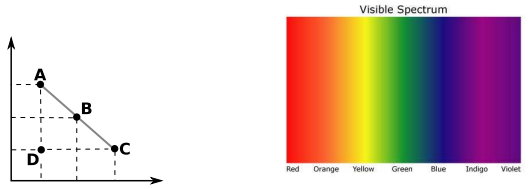
Morphing requires a description of the entities being morphed (shapes, images, sounds, etc), followed by establishing the correspondence between these descriptions, as depicted in figure 2b . The morph is achieved by a description whose elements are intermediate. Elements without correspondence in the description render the transformation complicated. If we are morphing faces and one of them has a mole, we will have to decide how we are going to represent intermediate versions of the unmatched feature. One of the consequences of the lack of correspondence betwen the objects being morphed is that we can have multiple possible transformations depending on how we decide to deal with the free feature. When morphing sounds, most works address the correspondence between model parameters [14, 11].

3.2. Intermediateness

The morphed objects should be perceived as intermediate. For example, when transforming between a square and a circle we want to avoid transforming the square into another recognizable shape first (say, a triangle) that is not perceived as intermediate between the square and the circle and then finishing the transformation from this shape into the circle. Intuitively, when transforming between a child's and a man's face, we expect all the hybrids to be human faces because it would be counterintuitive otherwise. Conceptually, the transformation should be the face of a person getting older.

Figure 3a illustrates the requirement of intermediateness. Points that are intermediate between A and B lie along the segment $\overline{AC}$ connecting the points, such that the distances from A to B plus the distance from B to C be equal to the distance from A to C. Notice that in figure 3a point D lies at the same distance from points A and C, yet, it is not intermediate between them.

In practice, when morphing shapes, images, and sounds, we make use of the interpolation principle as a convex combination [3, 4] of the parameters to achieve intermediateness. A convex combination leads to intermediate representations in the space of parameters. We argue that



(a) intermediateness (b) categorical perception

Figure 3: Depiction of intermediateness and categorical perception. Part (a) shows that point B is intermediate to points A and C, while point D is not. Part (b) illustrates categorical perception of colors by showing that a continuous variation of frequencies leads to a discrete perception of colors.

researchers must evaluate if the result is perceptually intermediate as well.

3.3. Smoothness

Ideally, we want to obtain morphed sounds that change gradually ("smoothly") from source to target. Since we control the transformation with the morphing factor α, continuously varying α should lead to a gradual transformation. When we suppose that perception of the stimulus is linear, adding the same factor should increase the perception by the same amount. In the case of morphing, most people intuitively expect a linear variation of the morphing factor α to produce the perception of a linearly gradual transformation. However, some stimuli are perceived continuously, others categorically due to their cognitive representation. When perception is categorical, even a continuous variation leads to a discontinuous percept that changes very little inside a category and much more between categories.

3.3.1. Categorical Perception

Categorical perception means that a change in some variable (stimulus) along a continuum is not perceived as gradual. Discrimination between stimuli is much more accurate between categories than within them. Figure 3b shows the visible spectrum of light with a continuous variation of the values of frequency. Our brains interpret this information categorically, (more or less) separated into stripes labeled red, blue, etc.

The perception of color stems from the cognitive representation of different wavelengths (or equivalently frequencies) of light. Color, on the other hand, is merely a cognitive label associated with certain socially constructed ranges of frequencies. When morphing sounds, we intrinsically assume that perception of the sound material used (musical instrument sounds, vocal utterances, environmental sounds, etc) is continuous. However, we should investigate if the morph is perceptually smooth. Here, an important theoretical issue plays a definitive role, the conceptual distance between the sounds being morphed.

3.3.2. Conceptual Distance

Let us suppose for a moment that the perception of faces is continuous. If we establish correspondence between the faces and interpolate parameters of a perceptually relevant representation (model), we should expect the morph to be smooth. However, some examples of face morph will lead to artificial hybrids simply because the faces are conceptually very far from each other. Figure 4 illustrates that the morph between a man's and a cat's face looks less natural than one between two human faces because of the conceptual distance between them. As a general rule of thumb, the naturalness of the morph is inversely proportional to the conceptual distance between them. The farther apart the objects are in the conceptual space, the more challenging it is to obtain convincing morphs.

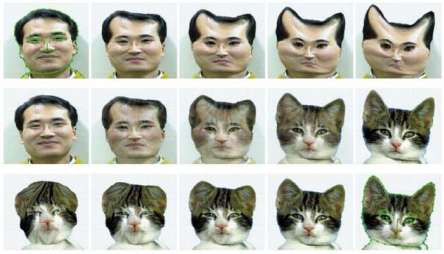


Figure 4: Effect of the conceptual distance on the morph.

4. EVALUATION PROCEDURE

Figure 5 depicts a flowchart with subjective and objective evaluation procedures for morphing. When the purpose of morphing is artistic, the evaluation of the results is usually subjective. Each individual will evaluate the results according to their own aesthetic criteria. When we want to obtain morphed results for technical applications, such as sound synthesizers or to study perceptual aspects of morphing (such as continuous timbre spaces), we need to establish an objective way of measuring the quality of the results. In figure 5 we see the three criteria proposed, namely correspondence, intermediateness, and smoothness. The objective evaluation procedure can use perceptual (listening tests) and automatic (feature values) means.

First of all, we must ensure correspondence between the representations of the sounds being morphed. Tellman [14] is among the first to investigate correspondence in sound morphing. The model used to represent the sounds plays a crucial role in this step [4, 3] because model parameter correspondence does not necessarily guarantee correspondence of perceptual features of sounds. Osaka [11] examines the problem of matching partials when interpolating the parameters of a sinusoidal model. Caetano [4], in turn, proposes a source-filter model to represent musical instrument sounds that guarantees temporal and spectral correspondence. More specifically, correspondence between the frames of the sounds and between spectral envelope parameters for each frame.

The perceptual intermediateness of the morph depends largely on how perceptually relevant the representation is. When interpolating the parameters of physical models, Hikichi [9] recognizes that the linear interpolation of the parameters does not lead to perceptually linear morphed sounds. So they propose to construct MDS spaces using the source, target and morphed sounds to study how to warp the interpolation factor to obtain perceptually linear morphed sounds. Naturally, this approach renders the results very difficult to obtain and to evaluate. Also, the warping function is model dependent and propably user dependent too, since it is subjective.

Finally, it is important to investigate the smoothness of the transition. Caetano [4, 3] proposes to investigate how accurately the morphing factor α controls the morph guided by perceptually salient features such as spectral centroid and attack time. Caetano varied the the morphing factor α linearly and studied which of several representations leads to linear variation of the feature values.

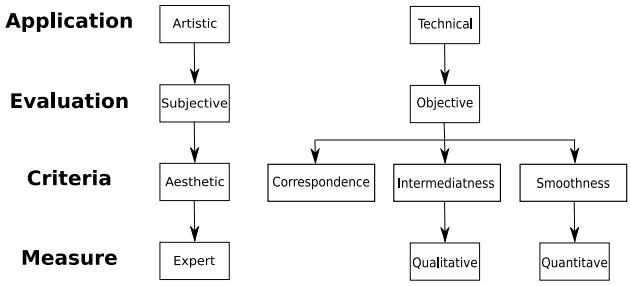


Figure 5: Subjective and objective evaluation procedure for morphing.

5. CONCLUSIONS AND FUTURE PERSPECTIVES

Most works about sound morphing skip the evaluation of the results because perceptual evaluations are cumbersome and expensive and there is no standard objective evaluation criteria or procedure when morphing. In this work, we proposed correspondence, intermediateness, and smoothness as evaluation criteria. We discussed each criterium in turn, giving examples to illustrate their importance. Then we presented an objective evaluation framework using the proposed criteria, along with perceptual assessments and quantitative measures. Future perspectives of this work include the development of a standard evaluation procedure for sound morphing that defines the perceptual tests and the feature values to investigate the correspondence, intermediateness, and smoothness of sound morphing algorithms.

6. ACKNOWLEDGEMENTS

This work is funded by the Marie Curie IAPP "AVID MODE" grant within the European Commissions FP7 and was partially carried out at IRCAM.

7. REFERENCES

[1] M. Ahmad, H. Hacihabiboglu, and A. Kondoz, "Morphing of transient sounds based on shift-invariant discrete wavelet transform and singular value decomposition," in *Proc. ICASSP*, 2009.

[2] A. Caclin, S. McAdams, B. K. Smith, and S. Winsberg, "Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones," *Journal Acoust. Soc. Am.*, vol. 118, no. 1, pp. 471–482, 2005.

[3] M. Caetano and X. Rodet, "Automatic timbral morphing of musical instrument sounds by high-level descriptors," in *Proc. ICMC*, 2010.

[4] ——, "Sound morphing by feature interpolation," in *Proc. ICASSP*, 2011.

[5] T. Ezzat, E. Meyers, J. Glass, and T. Poggio, "Morphing spectral envelopes using audio flow," in *Proc. ICASSP*, 2005.

[6] K. Fitz, L. Haken, S. Lefvert, C. Champion, and M. O'Donnel, "Cell-utes and flutter-tongued cats: Sound morphing using loris and the reassigned bandwidth-enhanced model," *Computer Music Journal*, vol. 27, no. 3, pp. 44–65, 2003.

[7] J. M. Grey and J. W. Gordon, "Perceptual effects of spectral modifications on musical timbres," *Journal Acoust. Soc. Am.*, vol. 63, no. 5, pp. 1493–1500, 1978.

[8] J. Harvey, "Mortuos Plango, Vivos Voco: A realization at ircam," *Computer Music Journal*, vol. 5, no. 4, pp. 22–24, 1981.

[9] T. Hikichi and N. Osaka, "Sound timbre interpolation based on physical modelling," *Acoust Sci Technol*, vol. 22, no. 2, pp. 101–111, 2001.

[10] M. McNabb, "Dreamsong: The composition," *Computer Music Journal*, vol. 5, no. 4, pp. 36–53, 1981.

[11] N. Osaka, "Timbre interpolation of sounds using a sinusoidal model," in *Proc. ICMC*, 1995.

[12] ——, "Timbre morphing and interpolation based on a sinusoidal model," in *Proc. ICA/ASA Joint Meeting*, 1998.

[13] M. Slaney, M. Covell, and B. Lassiter, "Automatic audio morphing," in *Proc. ICASSP*, 1996.

[14] E. Tellman, L. Haken, and B. Holloway, "Morphing between timbres with different numbers of features," *Journal Audio Engin. Soc.*, vol. 43, no. 9, pp. 678–689, 1995.