**KNOWLEDGECUDDLE PUBLICATION**

International Journal of Research in Advance Engineering (IJRAE) – Volume-1 Issue -1

# Animal Voice Morphing System

Siddhapara Milan v.
milansiddhapara@gmial.com
Gujarat Technological University,
Gujarat, India.

---

***Abstract:This paper describe the morphing concept in which we convert the voice of any person into pre -analyzed or pre-recorded voice of any animals.As the user generate a pre-established voice, his pitch, timbre, vibrato and articulation can be modified to resemble those of a pre-recorded and pre-analyzed voice of animal. This technique is based on SMS. Thus using this concept we can develop many funny application and we can used this type of application in mobile device, personal computer etc. for enjoying the sometime of period.***

***Keywords: Voice morphing, Animal.***

---

## I. Introduction

In general morphing means “conversion of some stages into some another stages”. Similarly in voice morphing us convert the voice generated from the any source into pre-analyzed or predict target voice in other word voice morphing means the transition of one speech signal into another. Like image morphing, speech morphing aims to preserve the shared characteristics of the starting and final signals, while generating a smooth transition between them. In voice morphing processone speech signal should smoothly change into another, keeping the shared characteristics of the starting and ending signals but smoothly changing the other properties. The major properties of concern as far as a speech signal is concerned are its pitch and envelope information. These two reside in a convolved form in a speech signal. Hence some efficient method for extracting each of these is necessary.The main goal of the developed voice morphing methods is the smooth transformation from one sound to another, thus, the combination of two sounds to create a new sound with an intermediate timbre. Most of these methods are based on the interpolation of sounds parameterizations resulting from analysis/synthesis techniques, such as the Short-timeFourier Transform (STFT), Linear Predictive Coding (LPC) or Sinusoidal Models.

In this paper we focus on specific concept, we convert the voice of any person into predict animal voice to generate some joyful environment. In this paper user convert the pitch, articulation and also able to control the degree of morphing to achieve the desire objective. Morphing take place in following stages, first system understand the voice of source (person) and find same sound Pieces from the target voice, then interpolate the selected attributes and produced the morphed voice In this paper [1] the implantation is discussed.

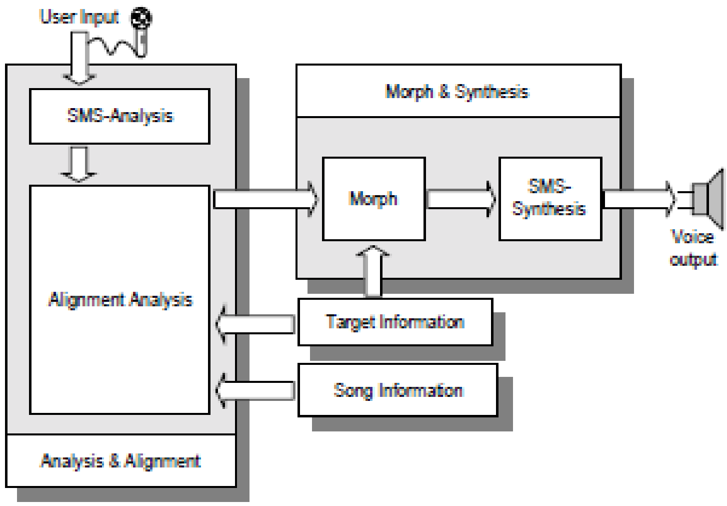


*Figure 1. System block diagram.*

## II. Voice Morphing System

Fig.1 shows the general block diagram of voice imitator system. The underlying analysis technique is SMS.Also in system recognition and alignment module was added for synchronizing the user's voice with the target voice before the morphing is done.Before we can morph a particular voice we have to supplyinformation about target information(predict animal voice ) and voice to be morphed. The system requires the phonetic transcription of the lyrics, the melody as MIDI data, and the actual recording to be used as the target audio data. Thus, a good impersonator of any animal that originally Voice has to be recorded. This recording has to be analysed with SMS, segmented into "morphing units", and each unit labelled with the appropriate note and phonetic information of the song. This preparation stage is done semi-automatically, using a non-real time application developed for this task. The first module of the running system includes the real-time analysis and the recognition/alignment steps. Each analysis frame, with the appropriate parameterization, is associated with the phoneme of a specific moment of the voice and thus with a target frame. The recognition/alignment algorithm is based on traditional speech recognition technology, that is, Hidden Markov Models (HMM) that were adapted to the singing voice [2]

The source sound are segmented into number of frames and each frame are individual analyzed. The recognition\alignment algorithm is based on hidden Markov models (HMM) [3] and when the source frame matched with target frame then we morphed them and get the desired output.

## III. Spectral Modeling Synthesis

We use the term SMS to describe our ongoing research on the development of techniques and software applications for the analysis, transformation and synthesis of musical sounds based on spectral models [2].Information on SMS and public domain software applications based on it can be found in our Web site (http://www.iua.upf.es/~sms).

## IV. Analysis Using SMS

Analysis processes can be categorized as follows.

- Preprocessing or representation conversion: This involves processes like signal acquisition in discrete form and windowing.
- Cepstral analysis or Pitch and Envelope analysis: This process will extract the pitch and formant information in the speech signal.
- Morphing which includes Warping and interpolation.

- Signal re-estimation.

The traditional SMS analysis output is a collection of frequency and amplitude values that represent the partials of the sound (sinusoidal component), and either filter coefficients with a gain value or spectral magnitudes and phases representing the residual sound (non sinusoidal component) [4]

The major improvement is done SMS model is alignment and analysis. A major improvement to SMS has been the real-time implementation of the whole analysis/synthesis process, with a processing latency of less than 30 milliseconds.

## V. Phonetic recognition/ alignment

This part of the system is responsible for recognizing the voice that is being generated from the userthat a similar segment can be chosen from the target information.

The huge amount of research is carry out in speech recognition.The recognition systems work well when tested in the well-controlled.The recognition systems work reasonably well when tested in the well-controlled environment of the laboratory. However, phoneme recognition rates decay miserably when the conditions are adverse.

### *A. SMorphing*

Speech morphing can be achieved by transforming the signal.To prepare the signal for the transformation, it is split into a number of 'frames' - sections of the waveform. The transformation is then applied to each frame of the signal. This provides another way of viewing the signal information. The new representation (said to be in the frequency domain) describes the average energy present at each frequency band. Further analysis enables two pieces of information to be obtained: pitch information and the overall envelope of the sound. A key element in the morphing is the manipulation of the pitch information. If two signals with different pitches were simply cross-faded it is highly likely that two separate sounds will be heard. This occurs because the signal will have two distinct pitches causing the auditory system to perceive two different objects. A successful morph must exhibit a smoothly changing pitch throughout. The pitch information of each sound is compared to provide the best match between the two signals' pitches. To do this match, the signals are stretched and compressed so that important sections of each signal match in time. The interpolation of the two sounds can then be performed which creates the intermediate sounds in the morph. The final stage is then to convert the frames back into a normal waveform.

However, after the morphing has been performed, the legacy of the earlier analysis becomes apparent. The conversion of the sound to a representation in which the pitch and spectral envelope can be separated loses some information. Therefore, this information has to be re-estimated for the morphed sound. This process obtains an acoustic waveform, which can then be stored or listened to.

Depending on the voice of person (user), a unit from the target is selected. Each frame from the user is morphed with a different frame from the target, advancing sequentially in time. Then the user has the choice to interpolate the different parameters extracted at the analysis stage, such as amplitude, fundamental frequency, spectral shape, residual signal, etc. In general the amplitude will not be interpolated, thus always using the amplitude from the user and the unvoiced phonemes will also not be morphed, thus always using the consonants from the user. This will give the user the feeling of being in control.In most cases the durations of the user and target voice to be morphed will be different. If a given user's voice is shorter than the one from the target the system will simply skip the remaining part of the target voice and go directly to the articulation portion. In the case when the user voiceisa longer than the one present in the target data the system enters in the loop mode. Each voiced phoneme of the target has a loop point frame, marked in the preprocessing, non-real time stage. The system uses this frame to loop-synthesis in case the user sings beyond that point in the phoneme. Once we reach this frame in the target, the rest of the frames of the user will be interpolated with that same frame until the user ends the phoneme.

## VI. CONCLUSION

The final purpose of the system is to make joyful voice from the user's (source) voice. Means using this we developed the many funny application but we need the some voice sample of target voice. For instance the voice of Karan is translated into voice of cat. But Limitation is that we need a clear voice from the source so we can easily recognize or analyzed it. If there is no clear voice of user then sample of that voice will converted into some another sound pieces of target animal. So if we have clear voice then we easily convert into target vice and enjoy the sometime of period.

## REFERENCES

[1]. Cano, P., A. Loscos, J. Bonada, M. de Boer, X. Serra.2000. "Singing Voice Impersonator Application for PC". *Proceedings of the ICMC 2000.*

[2]. Serra, X. 1997. "Musical Sound Modeling with Sinusoids plus Noise". G. D. Poli and others (eds.), *MusicalSignal Processing,* Swets & Zeitlinger Publishers, 1997.

[3]. Loscos, A., P. Cano, J. Bonada. 1999. "Low-Delay Singing Voice Alignment to text*". Proceedings ofthe ICMC* 1999.

[4]. Serra, X. 1997. "Musical Sound Modeling with Sinusoids plus Noise". G. D. Poli and others (eds.), Musical Signal Processing, Swets & Zeitlinger Publishers, 1997.