UNIVERSITI PUTRA MALAYSIA

MOTION PATH GENERATION USING A MODIFIED 6TH ORDER POLYNOMIAL FUNCTION FOR VISUAL SPEECH SYNTHESIS

SITI SALWA SALLEH

MOTION PATH GENERATION USING A MODIFIED 6TH ORDER POLYNOMIAL FUNCTION FOR VISUAL SPEECH SYNTHESIS

By

SITI SALWA SALLEH

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor Philosophy

February 2008

My special dedication is to…

My beloved husband, ***Aruddin bin Bakron*** *and my loving children,* ***Muhammad Amirul Hafizin*** *and* ***Amirah Syuhada*** *for their continous support, inspirations, motivation and being able to compromise in every way.*

May God Bless them all. Ameen.

Abstract thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirements for the degree of Doctor of Philosophy.

MOTION PATH GENERATION USING A MODIFIED 6TH ORDER POLYNOMIAL FUNCTION FOR VISUAL SPEECH SYNTHESIS

By

SITI SALWA SALLEH

February 2008

Chairman : Associate Professor Dr. Rahmita Wirza O.K. Rahmat

Faculty : Computer Science and Information Technology

Facial model that consist of synchronized speech and lips are able to increase speech intelligibility. This kind of system is called Visual Speech Synthesis system (VSS). Realisitic visual speech synthesis normally require manipulation of the facial mesh's vertices. These processes are complex; it requires large memory and computational power. Another technique that can be used for the same purpose is by using the parametric function which is able to control the motion of points on the lips model. Therefore, this study proposed the used of 6th order polynomial function as the lips' motion curve. The 6th order polynomial curve however is wild and unstable at the beginning and end of the curve.

It needs to be altered because it will be used as the motion curve. A formulation has been proposed in order to flatten the curvy portions. Subsequently the altered

polynomial curve is used to develop a computational steps that composes an isolated digit words utterance. This technique manages to generate the visual speech synthesis that start and end with neutral lips shapes. It also manages to increase the lips motion velocity and acceleration at the beginning of the utterance and decrease the motion velocity and acceleration when the utterance is complete. Another contribution of this study is the computational technique and steps to generate continuous utterance based on the altered 6^{th} order polynomial. This technique focused on lips motion in between one utterance to another. It also considers the motion velocity and acceleration in synthesizing continuous utterance. As a result it manages to produce realistic and smooth continuation. Synthesized visual speech was compared to the actual lips deformation to see the degree of its realistic realization. The actual motion curve is captured by using optical motion capture software. Motion similarity is measured base on the correlation coefficient values produce among the two curves. The control vertices relocation; lips width and height during speech; motion velocity and acceleration of control vertices and compare the shapes similarities between synthesize and actual lips were also measured. Results have shown that the use of 6^{th} order altered polynomial function is able to produce good speech synthesis with 88% - 95% similarity. In future the use of these techniques will improve in order to produce higher quality of visual speech synthesis.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan Ijazah Doktor Falsafah.

PENJANAAN PANDUAN PERGERAKAN MENGGUNAKAN FUNGSI POLINOMIAL TERTIB 6 YANG DIPINDA BAGI SINTESIS PERTUTURAN VISUAL

Oleh

SITI SALWA SALLEH

Febuari 2008

Pengerusi : Professor Madya Dr. Rahmita Wirza O.K. Rahmat

Fakulti : Sains Komputer dan Teknologi Maklumat

Model wajah beserta pergerakan bibir yang sinkroni dengan sebutan kata-kata telah terbukti berjaya meningkatkan keberkesanan komunikasi. Sistem applikasi yang mengandungi model wajah seperti ini dipanggil sebagai Sistem Sintesis Visual Pengucapan (SVP). Pada kebiasaannya penghasilan sintesis visual yang realistik memerlukan manipulasi titik-titik yang agak kompleks pada permukaan model wajah. Manipulasi seperti ini juga memerlukan ruang ingatan yang besar dan melibatkan kapasiti pengiraan yang tinggi. Alternatif kepada teknik manipulasi titik ialah dengan penggunaan fungsi parameter. Fungsi parametrik boleh digunakan bagi mengawal pergerakan titik kawalan pada model bibir. Di dalam kajian ini, fungsi parametrik polinomial tertib 6 digunakan sebagai panduan pergerakan model bibir. Akan tetapi lengkungan polinomial tertib 6 ini seringkali tidak terkawal dan berkedut pada permulaan dan akhir garis lengkungan. Memandangkan lengkungan

yang dihasilkan akan digunakan sebagai panduan pergerakan untuk sintesis model pertuturan, maka ianya perlu diperbaiki. Satu rumusan untuk meminda dua hujung lengkungan yang berkedut telah digunakan. Seterusnya lengkungan yang dipinda akan digunakan untuk menghasilkan sintesis visual pertuturan digit terasing yang dimulai dan diakhiri dengan bentuk bibir neutral. Teknik ini berjaya menghasilkan animasi pergerakan bibir yang meningkat halaju dan pecutannya pada permulaan sebutan dan menurun halaju dan pecutannya apabila sebutan berakhir. Bagi menghasilkan sintesis pertuturan berterusan, satu teknik lagi dibangunkan berdasarkan lengkungan polinomial tadi. Teknik ini memfokuskan kepada pergerakan model bibir diantara satu sebutan dengan sebutan seterusya. Ia mengambil kira halaju dan pecutan pertuturan apabila sebutan bertukar. Sintesis sebutan berterusan yang berkesinambungan dan pembentukan bibir yang realistik telah berjaya dihasilkan.

Kesemua sintesis visual yang dihasilkan kemudiannya diukur tahap keserupaannya. Perbandingan keserupaan diantara pertuturan sintesis visual dengan pergerakan bibir sebenar telah dibuat. Pergerakan bibir sebenar diperolehi dari perisian '*motion capture'*. Tahap persamaan diukur berdasarkan darjah korelasi pekali kedua-dua lengkungan tersebut. Pengukuran lebar dan tinggi bukaan model bibir juga dibuat terhadap pergerakan sintesis titik kawalan berbanding pergerakan titik fitur bibir sebenar. Selain itu, halaju dan pecutan titik kawalan juga dibandingkan. Berdasarkan penilaian yang dibuat, keputusan menunjukkan bahawa penggunaan lengkungan polinomial tertib 6 yang telah dipinda berjaya menghasilkan sintesis visual pertuturan yang baik dengan kejituan persamaan diantara lingkuan 88% -

95% . Kajian masa hadapan akan melibatkan pembaharuan terhadap penggunaan lengkungan polinomial tertib 6 sebagai panduan pergerakan sintesis visual pertuturan dengan menghasilkan teknik-teknik yang mampu menambah baik kualiti sintesis.

ACKNOWLEDGEMENT

First and foremost, all praises to Allah SWT for His blessing for I have gained the opportunity to complete this study.

My sincere gratitude to those who were involve in contributing their help, support and the time in making this study successful. It has been in my good fortune to have seeked the advice and guidance from many wise and knowledgeable people.

I would like to express my deepest appreciation and sincere thanks to my dedicated supervisor Dr. Rahmita Wirza O.K. Rahmat for her guidance, encouragements, comments, ideas and tolerances that led to a better quality thesis.

My special thanks to the supervisory committee members, Associate Professor Dr. Ramlan Mahmod and Associate Professor Dr. Hjh. Fatimah Dato' Ahmad for their guidance in helping me completing this study.

I also would like to thanks to the Universiti Teknologi MARA for awarding me the scholarship that make me able to carry such study. To all my friends who have joined me in this journey of knowledge, I wish them success and thanks in advance. Last but not least, to my beloved family, thank your for their scarifications.

SITI SALWA SALLEH

February 2008

I certify that an Examination Committee has met on 21st February 2008 to conduct the final examination of Siti Salwa Salleh on her Doctor of Philosophy thesis entitled "Motion Path Generation Using A Modified 6th Order Polynomial Function For Visual Speech Synthesis" in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the student be awarded the degree of Doctor of Philosophy.

Members of the Examination Committee were as follows:

Ali Mamat, PhD
Associate Professor
Faculty Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Abd Azim Abd Ghani, PhD
Associate Professor
Faculty Computer Science and Information Technology
Universiti Putra Malaysia
(Examiner)

Syamala C. Doraisamy, PhD
Faculty Computer Science and Information Technology
Universiti Putra Malaysia
(Examiner)

Aini Hussain, PhD
Professor
Faculty of Engineering
Universiti Kebangsaan Malaysia
(External Examiner)

HASANAH MOHD GHAZALI, PhD
Professor/ Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 26 May 2008

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirements for the degree of Doctor Of Philosophy. The members of the Supervisory Committee are as follows:

Rahmita Wirza O.K Rahmat, PhD
Associate Professor
Faculty Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Ramlan Mahmod, PhD
Associate Professor
Faculty Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Fatimah Ahmad, PhD
Associate Professor
Faculty Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

AINI IDERIS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 8 May 2008

DECLARATION

I declare that this thesis is my original work except for quotations and citations, which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

_____

SITI SALWA SALLEH

Date: 21 February 2008

LIST OF TABLES

LIST OF FIGURES

LIST OF GLOSSARY

Terms	Description
Articulatory	The act of giving utterance and expression.
Articulograph	A device utilizing alternating electromagnetic fields.
Arytenoid	A single muscle, arises from the posterior surface and lateral border of one arytenoid cartilage.
Bilabials	Sound that produced with both lips
Coarticulation	Changes in the articulation of a speech segment.
Cartilages	A translucent elastic tissue that composes most of the skeleton of vertebrate.
Consonant	A consonant is a sound made by a partial or complete closure of the vocal tract.
Ellipsoidal	A closed plane curve generated by a pointing moving in such a way that the sums of its distances from two fixed points is a constant.
Glottis	The elongated space between the vocal cords.
Labiodentals	Sound that utterances with the participation of the lip and teeth.
Laryngeal	The upper part of the trachea of a air breathing vertebrates that is humans.
Orofacial	Side view of the face.
Oscilloscope	An instrument in which the variations in a fluctuating electrical quantity appear temporarily as a visible wave form on the flurescent screen of a cathode-ray tube.
Spectral	A continuum of color form that resemble a color spectrum that consist of an ordered arrangement by a particular characteristic.

Velocity	The rate of change of position along a straight line with respect to time.
Viseme	Visual representation of mouth shape during utterance.
Vowel	A sound in spoken language that is characterized by an open configuration of the vocal tract so that there is no build-up of air pressure above the glottis.
Vocal	Having or exercising the power of producing voice speech.
Vocal folds	The lower pair of vocal cords each of which when drawn taut, approximated to the contralateral member of the pair and subjected to a flow of breath produces the voice.

TABLE OF CONTENT