



**UNIVERSITI PUTRA MALAYSIA**

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

**SITI SALWA SALLEH**

**FSKTM 2008 6**



**MOTION PATH GENERATION USING A MODIFIED 6<sup>TH</sup> 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 continuous 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 6<sup>TH</sup> 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). Realistic 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 6<sup>th</sup> order polynomial function as the lips' motion curve. The 6<sup>th</sup> 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<sup>th</sup> 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<sup>th</sup> 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**

**Februari 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 aplikasi 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 seterusnya. 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 21<sup>st</sup> February 2008 to conduct the final examination of Siti Salwa Salleh on her Doctor of Philosophy thesis entitled “Motion Path Generation Using A Modified 6<sup>th</sup> 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

Table		Page
2.1	Utterance Position of Consonant	37
2.2	Properties for Digit Words in Standard Malay	38
2.3	Phoneme Groups	39
3.1	Number of Iterations to Reduce Lips Vertices	60
3.2	An Example of Euclidean Distance for Utterance “SEMBILAN”	73
3.3	Training Dataset	77
3.4	Number of Frames for Each Utterance	79
3.5	States of Curve to Determine $\beta$ Value	87
3.6	Example of Factor Multiplication	89
3.7	Velocity Increment for Continuous Utterance.	97
4.1	Correlation Coefficient for Actual In-Between Transition Versus Synthesis Lips	129
4.2	Eigenlips for SM Digit	145
4.3	Similarity measurement of Lips Shapes for Isolated Utterance	146
4.4	Similarity Measurement of Lips Shapes for Continuous Utterance	147
4.5	Visual to Visual Speech Recognition Test Results	148

## LIST OF FIGURES

Figure		Page
2.1	The Fundamental Design of VSS Development	25
2.2	Frontal View of Major Facial Muscle for Speech Articulation	31
2.3	The Structure of Standard Malay Language Phonemes	37
2.4	Animation Path using Linear Curve	42
2.5	Animation Path Using Polynomial	42
2.6	Animation Path Using Hermite Spline curve	43
3.1	Frontal and Orofacial Views of Speaker	52
3.2	The Layout Design of Studio Set up	54
3.3	The Process of Acquiring the Lips Model	55
3.4	Parameters Used in Relevance Measurement Calculation	57
3.5	Vertices With Difference Relevance Measure	58
3.6	Example of Parameters On the Lips	59
3.7	Frontal View Of The Wire Frame Lips	61
3.8	Research Work Block Diagram	62
3.9	Detail Process of Research Work	63
3.10	Image Analysis Process	65
3.11	An Example of Neutral lips	65
3.12	An Example of Nostril and Binary Target Nostril	66
3.13	Lip Sequence for Word Utterance of “KOSONG”.	67

3.14	Example of Measurement Using Euclidean Distance	68
3.15	Position of Feature Points on The Lips Image	69
3.16	Reference Point of Lips in Grey Scale and Inverse Color Image	71
3.17	Distance Measurement Process and Computation	73
3.18	Flow of Process to Obtain Polynomial Function for Isolated Word Utterance	78
3.19(a)	X1 Vertex Motion Plot for Utterance “SEMBILAN”	81
3.19(b)	H1 Vertex Motion Plot for Utterance “SEMBILAN”	81
3.19(c)	H2 Vertex Motion Plot for Utterance “SEMBILAN”	82
3.20(a)	Motion Graph for Vertex X1 for utterance “SEMBILAN”	83
3.20(b)	Motion Graph for Vertex H1 for Utterance “SEMBILAN”	84
3.20(c)	Motion Graph for Vertex H2 for Utterance “SEMBILAN”	85
3.21	X1 Motion Curve for “KOSONG” Before and After Curve Modification	90
3.22	GAP Calculation	91
3.22	Proposed Pseudo-code to Generate Isolated Utterance Motion Curve	93
3.24	Modified X Motion Curve for Utterance “SEMBILAN” Before the Proposed Pseudo-code Implementation	94
3.25	Modified X Motion Curve for Utterance “SEMBILAN” After the Proposed Pseudo-code Implementation	94

3.26	GAP Calculation for Continuous Utterance	96
3.27	Pseudo-code to Compute Transition Curve for Continuous Utterance	98
3.28	Two Polynomial Function Stand Separately	99
3.29	Connection Establish In-Between Two Function After the Algorithm Implementation	99
3.30	Schematic Diagram of Visual to Visual Speech Recognition and Synthesis System	100
3.31	Steps in Comparing Lips Model Against DECface Visemes Dataset	102
3.32	Steps Involve in Histogram Intersection Computation	106
3.33	Visual to Visual Speech Recognition Procedure	107
4.1	The Correlation Coefficient Measurement of Lips Model Compared to DECface Dataset	112
4.2(a)	X1 Motion Curve for “SEMBILAN” Before and After Curve Modification	113
4.2(b)	Motion Curve H1 for “SEMBILAN” before and after Curve Modification	114
4.2(c)	Motion Curve H2 for “SEMBILAN” before and after Curve Modification	114
4.3(a)	The Comparison of X Motion Curve for Word Utterance “SEMBILAN”	117
4.3(b)	Correlation Coefficient of X Motion Curve for Word Utterance “SEMBILAN”	118
4.3(c)	The Mean and Standard Deviation of Difference of Synthesis and Actual	118
4.4(a)	The Comparison of H1 Motion Curve for Word utterance “SEMBILAN”	119
4.4(b)	Result for H1 Motion Curve for Word Utterance “SEMBILAN”	120



4.4(c)	The Mean and Standard deviation of Difference of Synthesis and Actual H1 Motion Curve for word Utterance “SEMBILAN”	120
4.5(a)	The Comparison of H2 Motion Curve for Word Utterance “SEMBILAN”	121
4.5(b)	Results for H2 Motion the Utterance “KOSONG”	122
4.5(c)	The Mean and Standard Deviation of Utterance of Synthesis and Actual H1 Motion Curve for Word Utterance “SEMBILAN”	122
4.6	Correlation Coefficient for Speaker 1 versus Synthesis Lips	124
4.7	Correlation Coefficient for Speaker 2 versus Synthesis Lips	125
4.8	Correlation Coefficient for Speaker 3 versus Synthesis Lips	126
4.9	Speaker 1 Versus Synthesis Lip Height for Continuous Utterance of Digit 0 to 9	131
4.10	Speaker 1 Versus Synthesis Lip Width for Continuous Utterance of Digit 0 to 9	131
4.11	The Mean and Standard Deviation Difference Between Synthesis and Actual Height of Lips for Continuous Utterance	132
4.12	The Mean and Standard Deviation Difference Between Synthesis and Actual Height of Lips for Continuous Utterance	132
4.13	The Comparison of Velocity of X Motion for Continuous Utterance of Digit Sequence 0 to 9	133
4.14	The Comparison of Acceleration of X Motion for Continuous Utterance of Digit Sequence 0 to 9	134
4.15	The Comparison of Velocity of H1 Motion for Continuous Utterance of Digit Sequence 0 to 9	134

4.16	The Comparison of Acceleration of H1 Motion for Continuous Utterance of Digit Sequence 0 to 9	135
4.17	The Comparison of Velocity of H2 Motion for Continuous Utterance of Digit Sequence of 0 to 9	135
4.18	The Comparison of Acceleration of H2 Motion for Continuous Utterance of Digit Sequence 0 to 9	136
4.19	The Comparison Of Lips Height For Sequence “SATU TIGA EMPAT LIMA TUJUH SEMBILAN “	137
4.20	The Comparison Of Lips Width For Sequence “SATU TIGA EMPAT LIMA TUJUH SEMBILAN “	138
4.21	The Mean And Standard Deviation of Difference Between Synthesis And Actual Width Of Lips	138
4.22	The Mean and Standard Deviation of Difference Between Synthesis and Actual Height of Lips	139
4.23	The Comparison of Velocity of X Motion for Continuous Utterance of Odd Digit Sequence 1 to 9	139
4.24	The Comparison of Velocity of H1 Motion for Continuous Utterance of Odd Digit Sequence 1 to 9	140
4.25	The Comparison of Velocity of H2 Motion for Continuous Utterance of Odd Digit Sequence 1 to 9	141
4.26	The Comparison of Acceleration of X Motion for Continuous Utterance of Odd Digit Sequence 1 to 9	141
4.28	The Comparison of Acceleration of H1 Motion for Continuous Utterance of Odd Digit Sequence 1 to 9	141
4.29	Steps to Identify Threshold Value	143

4.30	Eigenlips Image of Synthesis and Actual Lips Shape for Utterance “KOSONG”	145
4.31	Visual-to-Visual Speech Recognition and Error Rate	148

## LIST OF GLOSSARY

<b>Terms</b>	<b>Description</b>
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 <a href="#">arytenoid</a> 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 <a href="#">sound</a> in spoken <a href="#">language</a> that is characterized by an open configuration of the <a href="#">vocal tract</a> so that there is no build-up of air pressure above the <a href="#">glottis</a> .
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

	<b>Page</b>
<b>DEDICATION</b>	ii
<b>ABSTRACT</b>	iii
<b>ABSTRAK</b>	V
<b>ACKNOWLEDGEMENT</b>	viii
<b>APPROVAL</b>	ix
<b>DECLARATION</b>	xi
<b>LIST OF TABLES</b>	xvi
<b>LIST OF FIGURES</b>	xvii
<b>LIST OF GLOSSARY</b>	xxiii
 <b>CHAPTER</b>	
<b>1</b>	
<b>INTRODUCTION</b>	1
1.1 Background	1
1.2 Problem Statement	4
1.3 Objectives	7
1.4 Scope of Research	8
1.5 Research Methodology	10
1.6 Contribution of Research	11
1.7 Organization of Thesis	13
<b>2</b>	
<b>LITERATURE REVIEW</b>	15
2.1 Introduction	15
2.2 Review of Visual Speech Synthesis	19
2.2.1 Historical Background	20
2.2.2 Audio Visual Speech Application	20
2.3 Related Research on Parameter-based Audio Visual Speech Synthesis	21
2.4 The Fundamental Design of Visual Speech	25

	Synthesis Development	
	2.4.1 Visual Feature Analysis	26
	2.4.2 Visual Speech Animation Techniques	27
	2.4.3 Visual Synthesis and Implementation	29
2.5	Speech Production and Linguistic Aspect	29
	2.5.1 Lips Anatomy	30
	2.5.2 Speech Production Mechanism	31
	2.5.3 Speech Reading	32
	2.5.4 Visemes	33
	2.5.5 Standard Malay Language and Its Properties	34
2.6	Motion Planning in Animation	39
	2.6.1 Concept and Application	40
	2.6.2 Motion Path	41
2.7	Curve Fitting	44
	2.7.1 Polynomial Curve Fitting	45
	2.7.2 Least Square Fitting Method	46
2.8	Summary	48
<b>3</b>	<b>RESEARCH METHODOLOGY</b>	<b>50</b>
	3.1 Introduction	50
	3.2 Data Acquisition	51
	3.2.1 Speakers Profile	53
	3.2.2 Equipment Arrangement	53
	3.3 The 3D Lips Model	54
	3.4 Research Framework and Design	61
	3.5 Visual Feature Extraction	63
	3.6 Motion Planning and Path	74
	3.6.1 Polynomial Function Development	75
	3.6.2 Motion Path Generation for Continuous Utterance	90
	3.6.3 Motion Path Generation for Continuous Utterance	94

3.7	System Implementation: Visual to Visual Synthesis	100
3.8	Experimental Method and Set Up	101
3.8.1	Experiment to Measure Lips Deformation in Reduced Vertices Lips Model	101
3.8.2	Experiment to Compare the Effect on Before And After Polynomial Curve Modification	102
3.8.3	Experiment to Compare Synthesis Versus Actual Motion Of Isolated Utterance Visual Speech	103
3.8.4	Experiment to Compare Synthesized Versus Actual Motion Of Continuous Utterance Visual Speech	103
3.8.5	Experiment to Identify The Threshold Value for Lip Shape Similarity Measurement	104
3.8.6	Experiment to Measure Lips Shape Similarity	104
3.8.7	Experiment to Test Visual To Visual Speech Recognition	106
3.9	Summary	107
<b>4</b>	<b>RESULT AND DISCUSSION</b>	110
4.1	Introduction	110
4.2	Result From the Experiment to Measure Lips Deformation in Reduced Vertices Lips Model	111
4.3	Result from the Experiment to Compare Before and After Polynomial Fitting Alteration	112
4.4	Result from the Experiment to Compare Synthesis Versus Actual Motion of Isolated Utterance Visual Speech	115



4.5	Result from the Experiment to Compare Synthesized Versus Actual Motion of Continuous Utterance Visual Speech	127
4.5.1	Measurement on Utterance Transition	127
4.5.2	Measurement on Sequence of Utterances	130
4.6	Results form the Experiment to Identify The Threshold Value For Lip Shape Similarity Measurement	142
4.7	Results from the Experiment To Measure Lips Shape Similarity	144
4.8	Results from the Experiment on Visual to Visual Speech Recognition	147
4.9	Summary	149
<b>5</b>	<b>CONCLUSION AND FUTURE WORK</b>	<b>150</b>
5.1	Introduction	150
5.2	Performance of the Proposed Method	151
5.3	Suggestions for Future Work	152
	<b>REFERENCES</b>	<b>153</b>
	<b>APPENDICES</b>	<b>163</b>
	<b>BIODATA OF THE STUDENT</b>	<b>257</b>
	<b>LIST OF PUBLICATIONS</b>	<b>258</b>