A Speech Rate Related Lip Movement Model for Speech Animation

Zhou Wei, Wang Zengfu

Department of Automation, University of Science & Technology of China, Hefei, P.R.China
zhouwei8@mail.ustc.edu.cn  zfwang@ustc.edu.cn

Abstract

A novel lip movement model related to speech rate is proposed in this paper. The model is constructed based on the research results on the viscoelasticity of skin-muscle tissue and the quantitative relationship between lip muscle force and speech rate. In order to show the validity of the model, we have applied it to our Chinese speech animation system. The experimental results show that our system can synthesize the individualized speech animation with high naturalness at different speech rates.

Index Terms: speech rate, viscoelasticity, skin-muscle tissue, speech animation

1. Introduction

Realistic speech-driven facial animation is a heated issue in the field of Computer Graphics and has a lot of applications in human-computer interfaces, entertainment, perception research, telepresence, etc.

Since 1990s, facial animation has been developing rapidly. In 1993, K. Waters and T. Levergood developed a text-driven facial animation system DECFace, which used text as the input and converted the inputted text to speech with the synchronized animation [1]. In 1993, M. Cohen and D. Masarro put forward a co-articulation model for the purpose of synthetic visual speech [2]. In 1997, C. Bregler et al. constructed a context-dependent triphone visemes library and selected the optimal data from the library to match the phonemes sequence of the inputted audio track [3]. In 2000, S. Kshirsagar and N. Magnenat-Thalmann proposed a scheme of speech-driven facial animation, which converted natural speech information into facial animation parameters for synthesizing speech animation [4]. In 2003, Mingli Song et al. advanced a hybrid-driven approach, which treated both text and speech as inputs for more natural speech animation.

The above approaches can synthesize natural speech animation by adjusting facial animation parameters or constructing directly from high quality visemes and face databases in some restricted cases. However, these approaches seem not to be effective in synthesizing individualized animation. According to Kuehn & Moll and Ostry & Munhall’s researches, in continuous speech, speech rate has a strong effect on the velocity and amplitude of lip movement. At different speech rates, different people select different strategies of lip movement. For increased rate, some speakers decrease amplitude but maintain the velocity of the movement; others increase velocity while maintaining amplitude; and others make adjustments in both parameters [6][7]. Regarding the above relative researches as reference, this paper presents a new speech rate related lip movement model. Firstly, we find the relation between lip muscle force and speech rate by analyzing EMG signals with different speech rates. Secondly, we conclude and model the relation between lip movement velocity and amplitude according to the viscoelasticity of skin-muscle tissue. Finally, we apply the model to our speech-driven facial animation system and design the speech rate related speech animation.

The rest of this paper is organized as follows. In Section 2, we review the relation between muscle force and speech rate. In Section 3, the speech rate related lip movement model is presented. Section 4 shows some experimental results of the proposed approach. Section 5 concludes the paper with some suggestions for future work.

2. Relation between muscle force and speech rate

The electromyography (EMG) represents the excitation degree of muscle activities. According to [8][9], there exists a closed relation between EMG signal and speech rate.

Carlo J. De Luca analyzed the quantitative relationship between forces and the corresponding EMG signals of three different muscles and acquired the normalized Force/EMG Signal relationship curves [8]. According to the analysis of the curves, we can conclude approximately that EMG signal amplitude increases linearly for increased muscle force. It can be shown in Equation (1) as follows,

\[ F = aA + b \]  

where, \( F \) refers to the muscle force, which is considered as a constant during contraction, \( a, b \) are two constants, and \( A \) refers to the amplitude of EMG signal.

Amy B. Wohlert and Vicki L. Hammen measured the surface EMG signal from 20 adults who read a paragraph at different speech rates and acquired EMG Signal Amplitude/Speech Rate relationship curves [9]. The measurement results indicate that, the higher the speech rate is, the larger the EMG signal amplitude is. We can conclude approximately that the EMG signal amplitude of lower lip increases linearly while the amplitude of upper one increases segmentally linearly. It can be shown in Equation (2) as follows,

\[
\begin{align*}
A_{upper} &= cR + d & \text{if } R_1 \leq R < R_2 \\
A_{upper} &= eR + f & \text{if } R_2 \leq R \leq R_3 \\
A_{lower} &= gR + h & \text{if } R_4 \leq R \leq R_5
\end{align*}
\]  

where, \( A_{upper} \) and \( A_{lower} \) refer to the EMG signal amplitudes of upper lip and lower lip respectively, and \( R \) is the speech rate. The three threshold speech rates \( R_1, R_2, \) and \( R_5 \) are acquired from the EMG Signal Amplitude/Speech Rate relationship curves, which define the different intervals in segmental linear curves. From Equation (1) and (2), we can acquire Equation (3) as follows,

\[
\begin{align*}
F_{upper} &= C_1R + D_1 & \text{if } R_1 \leq R < R_2 \\
F_{upper} &= C_2R + D_2 & \text{if } R_2 \leq R \leq R_3 \\
F_{lower} &= C_3R + D_3 & \text{if } R_4 \leq R \leq R_5
\end{align*}
\]
where, \( C_1 \sim C_3, D_1 \sim D_3 \) are all constants; and \( F_{\text{upper}} \) and \( F_{\text{lower}} \) are the muscle forces of upper and lower lips, respectively. Therefore, the relationship between the muscle force of lip and speech rate can be expressed by \( F = f(R) \).

3. Speech rate related lip movement model

The study on biomechanics of the skin reveals that the skin is viscoelastic in its response to stress and strain [10]. As shown in Figure 1, in the lip area, there are a lot of linear muscles connecting the bone and the lip skin tissue which can control the lip movement. Figure 2 illustrates the points of the skin that adhere to the muscles in the lip skin-muscle tissue area. Therefore, the area which covers lip muscle can be considered as an independent viscoelastic system. A simplified model of the viscoelastic system is shown as Figure 3:

![Figure 1: Muscles in the lip area connecting the bone and the lip skin tissue](image1)

![Figure 2: The points of the skin that adhere to the muscles in the lip skin-muscle tissue area](image2)

![Figure 3: The model of viscoelastic system, where \( k \) is the basic spring stiffness, \( c \) is the viscid damp coefficient](image3)

In the process of muscle contraction, muscle force can be considered as a step excitation in the viscoelastic system. The dynamics equation of a point of the skin that adheres to the muscle in the lip skin-muscle tissue area is shown in Equation (4) as follows,

\[
m x + c x + k x = F(t)
\]

where, \( x \) refers to the point’s movement displacement, \( F(t) = F_0(\omega) \) refers to the muscle force, \( u(t) \) is called a unit step function, and \( F \) refers to the amplitude of the muscle force. It is supposed that in the process of constant muscle contraction, \( k \) can be considered as invariable; in the meanwhile, in the rapid muscle contraction process, the skin-muscle tissue can be simplified into an underdamping system with invariable damping ratio. The response to the step excitation \( F(t) \) is shown in Equation (5) below,

\[
x(t) = \frac{F}{k} \left[ 1 - e^{-\xi \omega_d t} \cos(\omega_d t - \varphi) \right]
\]

where

\[
\varphi = \arctan \frac{\xi}{\sqrt{1 - \xi^2}}
\]

\[
\omega_d = \omega_n \sqrt{1 - \xi^2}
\]

\[
\omega_n = \sqrt{\frac{k}{m}}
\]

In underdamping system, the vibration frequency of muscle is

\[
T = 2\pi \sqrt{\frac{m}{k(1 - \xi^2)}}
\]

where, \( \xi \) refers to the damping ratio. The muscle contraction time is

\[
t_{\text{con}} = \frac{T}{4} = \frac{d}{v}
\]

where, \( d \) refers to the displacement of a point of the skin that adheres to the muscle in one of the lip skin-muscle tissue areas, which moves from natural state to target position. \( v \) refers to the point’s average velocity during the muscle contraction. The Equation (8) can be concluded from Equation (6) and (7) as follows,

\[
k = \frac{\pi^2 m v^2}{4(1 - \xi^2) d^2}
\]

In terms of muscle activity, stiffness of skin-muscle tissue is the ratio of contraction force to change in muscle length [11] as is shown in Equation (9),

\[
K = \frac{F(t)}{l}
\]

where, \( l \) refers to the change of muscle length, which can be considered to be equal to \( d \) approximately.

According to the research result of Yu Zhang et al. [12], the stiffness function of skin-muscle tissue is given, showing in Equation (10),
\[ K = k(1 + d^2)^\alpha \]  

In (10), \(\alpha\) is the nonlinearity factor controlling the modulation. Function \(K\) can model different stress-strain relationship with different values of \(\alpha\).

According to the equations above and the relationship between muscle force and speech rate, which is expressed in Section 2, the speech rate related lip movement model is presented by giving the functions for \(v\) and \(d\) as follows (\(\alpha = 0.5\)):

\[ v = \frac{4d(1 - \xi^2) f(R)}{\pi^2 m \sqrt{1 + d^2}} \]  

\[ d = \frac{\pi^4 m^2 v^4}{16 f^2(R)(1 - \xi^2)^2 - \pi^4 m^2 v^4} \]

This model, which synthesizes viscoelasticity model of skin-muscle tissue and the relationship between muscle force and speech rate, simulates the displacement of the point of the skin that adheres to the muscle in one of the lip skin-muscle tissue areas and the average velocity during the muscle’s contraction. For the viscoelasticity of skin-muscle tissue, the other points’ displacements and velocity are dependent to \(d\), \(v\) and the positions in the influenced regions. Therefore, from the model we can conclude that:

1) At the same speech rate, the lip movement velocity increases with the lip movement amplitude.

2) At the same lip movement amplitude, the lip velocity increases with the speech rate.

3) At the same speech rate, the lip movement amplitude increases with the lip movement velocity.

4) At the same lip movement velocity, the lip movement amplitude deceases with the speech rate.

4. Experimental results

In our previous work, the lip muscle model [13] and the visual co-articulation model for Chinese mandarin [14] have been constructed. Due to the space limitations, the detailed presentation is ignored here. Based on Waters’ muscle model [15], our presented lip muscle model, which simulates the muscle movements by defining several muscle submodels and muscles’ influenced regions in lip area, can synthesize the lip shapes more accurately. All necessary lip shapes are synthesized by adjusting muscle parameters and the combination of all possible submodels. According to the rules of Chinese mandarin co-articulation and the context-dependent triphone model in continuous speech, we have constructed the triphone model-based visual co-articulation on the basis of the muscle model and then synthesized all the possible phone-viseme mappings thus produced the speech animation in habitual rate.

In order to synthesize the speech animation with different individuated lip movement strategy at different speech rates, we improve the lip movement model of our implemented speech animation system by adding two parameters — speech rate \(R\) and lip movement average velocity \(v\), which are acquired from the recorded sound track and the video sequence. We obtain different kinds of lip movements at different speech rates through following steps:

Firstly, the average velocity of the point in the lip muscle area is estimated through analyzing the human lip shape video sequence with optical flow estimation approach [16]. An example of the flow estimates of the lip for the transition between Figure 4(a) and Figure 4(b) can be seen in Figure 4(c), where flow vectors have been overlaid on the lip image. Secondly, the lip muscle force can be computed through the proposed function \(F = f(R)\) by using the acquired speech rate and EMG signal amplitude in the lip muscle region. Finally, according to Equation (12), from its natural state to the target position, the displacement of the point of the skin that adheres to the muscle is obtained through a certain lip movement strategy. The direction of \(d\) is the same as the muscle force. Other points’ displacement in the lip muscle’s influenced area can be computed by our proposed lip muscle model. Therefore, based on the acquired displacement of the lip movement, all possible phones’ corresponding key lip shapes, which are at different speech rates, can be synthesized by adjusting lip muscle parameters.

In our experiment, the speaker read a paragraph at habitual, fast and slow speech rates respectively through the lip movement strategy of changing the lip amplitude but maintaining the velocity of the lip movement. The habitual, fast, and slow speech rate are separately 190, 286, and 117 words/min. Setting “Zhong Guo Ke Da” for example, the obtained lip shapes video sequences at the three speech rates are shown in the upper rows of Figure 5-7; and the correspondent synthesized lip shapes sequences at the three speech rates are shown in the lower rows of Figure 5-7.

![Figure 4: Optical flow estimation images. (a) and (b) are two adjacent phones’ corresponding lip shapes. (c) illustrates the flow estimates of the lip for the transition between (a) and (b).](image)

![Figure 5: Lip shapes sequence at habitual speech rate (190 words/min)](image)

![Figure 6: Lip shapes sequence at fast speech rate (286 words/min)](image)
5. Conclusion and future work

A novel lip movement model related to speech rate is proposed in this paper. The model is constructed by the analysis of the relationship between muscle force and speech rate and the research on the viscoelasticity of skin-muscle tissue. The speech animation at different speech rates is produced by using the proposed model on a basis of our previous work.

The approach put forward in this paper analyzes the relationship between the speech rate and lip movement. Compared with the previous model, this proposed lip model can simulate different lip movement strategy of different people with high degree of individuality.

At the same time, there are some promising research directions in our future work. Our proposed lip movement model is based on a lot of independent viscoelastic system. In this system, we only focus on the viscoelasticity of each lip muscle area, which does not accord with the multi-degree mass-spring network model of the real skin-muscle tissue. Therefore our proposed model may be lacking in the ability of displaying realistic details. We are going to construct a lip model by using mass-spring network, though it is complicated and of great challenge.

Besides the speech rate, speech volume, speech tone and emotion et al. can affect the lip activity. We should make further research into these factors and take them into account when we try to improve the lip model, which will help perfect our work.

6. References


