A sign language smart phone based on realistic animation

Wentong Wang*, Wenhui Zhang, and Bin Sun

Beijing University of Technology, Beijing 100124, China

Abstract. Mobile phones are now popular in current society for its portability and intelligence, but excluded from life of the hearing-impaired because sounds are unavailable for them. This paper provides a framework of speech based Chinese Sign Language (CSL) Animation System on Smart Phones, which combines speech recognition and sign language animation to make phones convenient for the hearing-impaired. This paper recognizes the Chinese speech into text first, and then proposes a HMM-SVM model to judge interrogative expression from the speech and rewrites the plain text as mark-up text according to the rules of Chinese Sign Language Mark-up Language (CSLML) which provides signs for recording expressive context in the text, the corresponding animation is finally synthesized using key-frame technology. The prototype system based on the framework provided in this paper gives plausible Sign Language animation.

1 Introduction

A sign language (also named as signed language or simply signing) is a language that uses manual communication and body language to convey meaning, instead of acoustically conveyed sound patterns. This can involve simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions to fluidly express a speaker's thoughts [1]. However, in the common society, spoken languages, which depend primarily on sound, are the main carrier through the media or long-distance communication, which makes an enormous barrier to the deaf people from obtaining information and getting along with hearing people. According to the data from World Federation of the Deaf, there are around 70 million people with hearing deficiencies all over the world. However, the amount of sign language interpreters is far from what we really need. In the US, the ratio of the deaf and interpreters is 93 to 1. Finland has the best ratio-6:1, and Slovakia, the worst, is 3000:1 [2].

In China, according to the China Disabled Persons' Federation website, China has approximately 21 million people with hearing loss [3]. However, China urgently needs to build a system of training, testing and authentication of sign language interpreters. Sign language interpreting as a career, unlike the western countries, is still in the fledging stage. The insufficient of professional interpreters is far away from meeting the demand of the market.

* Corresponding author: wwt@emails.bjut.edu.cn

As the development of CSL animation system, which enables a TTSL (text to sign language) procedure, is partially playing an interpreting role on the PC platform [4] and Set Top Box (STB) system [5]. However, the input method could be a tricky issue. The programs on the PC and STB need users to prepare the transcription beforehand, which seems impossible to accomplish on cell phones. In this paper, we propose a combination of voice recognition system to make this up. Users can get transcripts and corresponding speech simultaneously from the incoming call, which will fulfil deaf people’s expectation of communicating through phones.

2 Speech based Chinese Sign Language Animation System

In this paper, we propose a Chinese speech to sign language animation method for Android phones and design it support deaf people to answer phone calls from arbitrary phones (not only from smart phones).

Figure 1 shows the module diagram of this system. By receiving the phone call, first, we use Google voice recognizer Intent [6] (an API to translate speech to text) to acquire the transcript corresponding to the input speech series. After obtaining the texts, we employ a text analyse step to recognize the interrogation of the speech. Then we have a set of interrogative expression marked texts. With CSLML scheme, we utilize the marked texts to synthesis CSL animation with interrogative expression. Our method in prototype enables CSL animation with plausible facial expression from speech through cell phone. Due to the limit of CSL recognition technique, there is only one direction in our prototype; it will, however, greatly facilitate the deaf using phones to communicate with other people from any phones from remote.

Fig. 1. Framework of speech based Chinese Sign Language (CSL) Animation System.

3 Phone Call to Interrogative Intonation Marked Texts

In this paper, we use a Google Intent [6] - voice recognizer Intent (an API component to translate speech to text). This API is pre-installed on most android phones, so we can use this Intent easily.

We create an Android Service to listen to the status of the phone. If a phone call comes, and is off hook, our app will start the speech recognizer service to generate transcriptions for the interrogation filter, as is shown in figure 2.

Interrogative mood is very important in communications [7-8]. The sentence with an interrogative mood could be a very different meaning. In this paper, after getting the transcript, we tried to extract interrogative semantics of the incoming speech.

While google voice recognizer does not provide intonational results, we utilize an interrogation filter to mark the interrogative tone. The interrogation filter contains two

phases: First, we check interrogative pronouns in transcript. The second phase is an HMM-SVM classification model to determine whether the intonation is increasing. After these two phases we can get an interrogative marked CSLML, which can be further utilized in the CSL synthesis system. Pronoun checking is fast and HMM-SVM classifier is a supplement.

Fig. 2. Voice to CSLML.

Fig. 3. CSLML to CSL animation with prosody.

3.1 Pronoun checking

Interrogative pronouns and modal particles are primary symbols of interrogative sentences. If a sentence contains an interrogative pronoun or interrogative modal particle, the mood of the sentence is interrogative. Pronoun checking is just a word to word matching. It is time-effective.

The interrogative pronouns include who (*shei*), what (*shenme*), and where (*nali*) and modal particles include *ma*, *ne*, *ba* (Chinese modal particles).

In this phase, first we split the sentence into words utilizing **jieba** – a word segmentation utility. Second, we search interrogative pronouns or modal particles in the word list. If there exists, we skip the HMM-SVM phase and mark the sentence as interrogative sentence. If not found, we then get into the second phase, HMM-SVM model.

3.2 HMM-SVM model

Interrogative pronouns and modal particles are primary symbols of interrogative sentences. If a sentence contains an interrogative pronoun or interrogative modal particle, the mood of the sentence is interrogative. Pronoun checking is just a word to word matching. It is time-effective.

First, we transform the incoming speech into MFCC features, and feed the MFCC features into a HMM model to extract temporal features. We then feed the temporal features into a SVM classifier, to determine whether the speech is interrogative.

4 Text to CSL Animation with Interrogative Expression

With the CSLML interface [9], we separate the CSLML texts into manual gesture part and non-manual gesture part. Manual gesture part is the transcription and the non-manual gesture part is expression information. Then we separately look up the manual gesture and non-manual gesture in corresponding database for corresponding data [10]. After attaining the data, our synthesis system displays the corresponding data frame by frame. The workflow is shown in figure 3. It's the first time using avatar for sign language on smart phones.

Facial expression is used in sign languages to convey specific meanings. In American Sign Language (ASL), for instance, raised eyebrows combined with a slightly forward head tilt indicate that what is being signed is a yes/no question. Lowered eyebrows are used for wh-word questions [11-12]. While in Chinese Sign Language, interrogative mood is expressed by only raising eyebrows [3]. The synthetic result is shown in figure 4~7.

5 Experimental results and analysis

5.1 Dataset

To train the HMM-SVM model, we collected a dataset with 10 interrogative speech sentences and 10 non-interrogative speech sentences. We employed 10 performers (5 males and 5 females). Each person records 20 sentences and make up this dataset. The sentences are daily conversations.

5.2 Training and Testing Results

We split the dataset with 80% for training and 20% for testing. We select RBF kernel in SVM model. From table 1, we can get pretty good accuracy for interrogative recognition. With pronoun-checking, we can only find sentences with apparent interrogative words. However some Chinese interrogative sentences do not use evident interrogative words in oral Chinese, such as *Zhe shi ni de ping guo?* ('Is it your apple?'). In this sentence, the speaker abbreviated the modal particle, but with an interrogative intonation, it can still express a questioning meaning. With HMM-SVM model, speech sentences with apparent interrogative intonation could easily be detected, but some interrogative samples in the database are in plain intonation, which cannot be classified correctly. Thus, combine this two method, could make a compliment, and greatly improved recognition accuracy.

The interpreting results relies on the speech recognition results. If the speech recognition is correct, we can output the correct CSL animation. The synthetic result is shown in figure 4~7. The same Chinese words is *nihao*. With neutral mood, the sentence means "Hello". However, with interrogative mood, it means "Are you OK?". Synthetic results show the differences between neutral and interrogative expressions in Chinese Sign Language.

Table 1. Interrogative recognition result.

Accuracy	w/o HMM-SVM	w/ HMM-SVM
w/o pronoun-checking	-	85%
w/ pronoun-checking	70%	90%

6 Conclusions

In this paper, we propose a Chinese speech to sign language animation model for Android phones and design it support deaf people to answer phone calls from any phones (not only smart phones). To make the sign language more reliable and realistic, we enhanced the interrogative expressions with an interrogative pronoun-checking procedure a and HMM-SVM model. Synthetic results show the differences between neutral and interrogative expressions in Chinese Sign Language. As facial expression is very important in Chinese Sign Language expression, in the future, we will introduce more expression in the system.

Fig. 4. Keyframes of “nihao”(‘Hello’).

Fig. 5. Keyframes of “nihao?” (‘Are you OK?’).

Fig. 6. Facial details of “nihao”(‘Hello’).

Fig. 7. Facial details of “nihao?” (‘Are you OK?’).

References

1. Sign language. In Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Sign_language.

2. V. López-Ludeña, R. Barra-Chicote, S. Lutfi, J. M. Montero , R. San-Segundo. LSESpeak: A spoken language generator for Deaf people. *Expert Systems with Applications*. (2012)
3. China Disabled Persons' Federation. <http://www.cdpf.org.cn/>
4. Z. Wang, W. Gao. A Method to Synthesize Chinese Sign Language Based on Virtual Human Technologies (in Chinese). *Journal of Software*, 10. (2002)
5. S. Wang. Chinese sign language synthesis driven by speech and text [M]. Beijing: Beijing University of Technology, (2011)
6. Google Intent Component. <https://developer.android.google.cn/reference/android/content/Intent.html>
7. C. N. Li, S. A. Thompson. Mandarin Chinese: A functional reference grammar. University of California Pr. (1989)
8. G. N. Leech. English in Advertising: Linguistic Study of Advertising in Great Britain. Longmans, Green & Company. (1966)
9. K. Ye, B. Yin, L. Wang. CSLML: a markup language for expressive Chinese sign language synthesis. *Computer Animation and Virtual Worlds*, 20(2-3), 237-245. (2009)
10. M. Delorme, M. Filhol, A. Braffort. Animation generation process for sign language synthesis. In *Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences on* (pp. 386-390). IEEE. (2009)
11. S. Dachkovsky & W. Sandler. Visual intonation in the prosody of a sign language. *Language and Speech*, 52(2-3), 287-314. (2009)
12. C. Baker, and D. Cokely. American Sign Language: A teacher's resource text on grammar and culture. Silver Spring, MD: T.J. Publishers. (1980)