

# Application of a Sign Language Synthesis System in Digital Library Services

Jing WAN, Bin WU, Yingxin QIAO

*Department of Computer and Network Systems, National Library of China, Beijing, 100081*

**Abstract:** Information Accessibility for disabled people is one of the most important design criteria for the China National Digital Library (CNDL) development. Sign language synthesis systems are effective to provide information services for people with hearing and speech impairments. This paper presents a framework of a sign language synthesis system application in CNDL, as well as discusses the relevant technologies applied in the system. CNDL has been a real practice area for the sign language synthesis research.

**Keywords:** sign language synthesis, digital library, on-line seminar, sign language presentation

## 淺析手語系統在數字圖書館中的應用

萬 靜, 吳 斌, 喬穎欣

國家圖書館計算機與網絡系統部 北京 100081

**摘 要:** 為殘疾人服務是國家數字圖書館建設的一個重要組成部分。其中，手語合成系統的應用將為聽力和語音障礙者（聾啞讀者）瀏覽數字圖書館資源（特別是在線講座等）提供信息無障礙的服務。本文論述了該手語合成系統的總體架構和採用的相關技術。國家數字圖書館的在線講座服務將是手語合成系統的一個重要的應用領域和實踐。

**關鍵詞:** 手語合成，數字圖書館，講座，三維手語顯示

### 1. 背景簡介

為廣大殘疾人服務是國家數位圖書館建設的一個重要組成部分。

通常，傳統圖書館服務所面對的殘疾人群體可以分為肢體殘疾（多指輪椅使用者）、視覺殘疾（盲人）和語音聽覺殘疾（聾啞人）三類。由於數位圖書館的許多內容都是通過互聯網和個人電腦來提供閱覽，對於肢體殘疾的讀者來說，在電腦操作和網路使用上，應該同健全讀者無異，而數位圖書館“足不出戶”的使用特性，正好為他們提供了巨大便利。因此，相對於傳統圖書館面向服務的三類殘疾人群，視覺殘疾（盲人）和語音聽覺殘疾（聾啞人）兩類殘疾人群則是數字圖書館的要重點

關注的物件，需要配以特殊的技術輔助。

對於盲人讀者，國家數字圖書館已與 2008 年 10 月開通了“中國盲人數位圖書館”的門戶網站，採用了多種比較成熟的資訊無障礙技術，本著公益服務、資訊讀取無障礙、邊建設邊服務、資源特色化與共建共用的原則，提供了大量的數位圖書和音頻數位資源，服務于盲人。同時，社會上也有很多成熟的讀屏軟體可供盲人在使用電腦時理解螢幕資訊。

相比而言，可供聾啞人使用的資訊無障礙產品則相對匱乏，許多音視頻資源無法供聾啞人使用。雖然現在大多數的解決方案是提供相應的字幕，但聾啞人對可視文字的理解速度要比手語慢得多。經研究表

明，聾啞人對手語的理解速度和健全人對語音語言的理解速度是相當的，手語可以說是聾啞人的“聽力系統”[1]。手語更直觀更迅捷，它是聾啞人最重要的交流手段[2]。

國內一些學者對聾啞人適用的資訊無障礙的相關技術，比如視頻編碼與分析、手語識別與合成、人臉識別等，正在做一些研究，並取得了一定的成果。但這些技術成果還處於實驗室階段，缺乏具體的應用領域[3-5]。正在建設中的“國家數字圖書館”為此提供了一個應用平臺。目前，國家圖書館正在進行如何將手語合成技術應用到國家數位圖書館服務領域的研究和實踐。本文將簡要介紹這一手語合成系統的構建思路和技術實現。

## 2. 手語合成系統在數位圖書館中的應用

正在建設中的國家數字圖書館，遵循“邊建設便服務”的原則，至今已有豐富的數位資源上線服務。對於

電子圖書、古代文獻、圖片等“無聲”的數位資源，聽力和語音障礙者（聾啞人）在使用上沒有障礙；而對於視頻影像資源，聽力和語音障礙者就有障礙，需要經過“字幕”等加工處理才能讓他們理解。

國家圖書館有大量的講座視頻資源，比如“文津講壇”、“文津圖書沙龍”等已成為國圖的講座品牌，幾乎每週都有講座舉辦，並連續3年將國內外著名專家學者的精華講座製作成多媒體數位檔並發佈，現在已經有387場、近100GB的數字講座得到授權，在互聯網上發佈。這是國家數位圖書館數位資源的重要組成部分。然而對於聾啞人讀者來說，目前講座現場不提供字幕和手語翻譯支援，在講座錄製加工後，仍有一部份講座沒有字幕，這使得聽力和語音障礙者無法欣賞講座的精彩內容，更無法親身參與到現場的互動活動。

為了改善對聾啞殘疾人讀者的服務，更好地幫助他們理解到講座的內容，我們希望通過採用手語合成的技術，對視頻講座配以手語畫面（效果圖如圖1），



圖1. 三維手語示例效果圖（主畫面是國家圖書館名譽館長任繼愈先生的講座《中國哲學的未來》的視頻，右上角的三維模擬人通過手語對講座的內容進行同步翻譯）。

Figure 1. A 3D effect of a sign language demonstration (the main picture is a lecture video of the honorary director of the National Library of China Ren Jiyu on “the future of Chinese philosophy”, the right corner above shows a 3D dummy man interpreting the lecture simultaneously).

無論聾啞人讀者即時參與講座還是後來觀看網路數位錄影，都能夠快速地理解講座內容，和正常人一樣，真正做到資訊無障礙。手語合成系統在國圖講座中的應用，將是國家數位圖書館為聾啞人提供服務的一個重要的應用領域，將極大地提高他們獲取資訊的自立能力。從理論上講，該手語合成系統，可擴展應用到數位圖書館的任何凡有交流性質的服務中，為廣大聾啞讀者提供便利。

### 3. 手語合成系統的構建

#### 3.1 總體流程

手語合成系統的總體流程是系統根據語音輸入（即時或多媒體影像）進行語音識，轉換成文本資訊，然後自動把文本資訊翻譯成對應的手語資訊，並採用圖像技術，把手語逼真地展示出來。

手語合成系統的總體流程如圖 2，包括語音識別、文本到手語碼的轉換、手語運動資料的建立、手

語動作的合成、三維手語的顯示等主要環節。

語音信號的來源可以是現場麥克風採集的聲音，或者現場錄製成多媒體資源中的音頻。音頻信號經過語音識別轉化為文本資訊，對這些文本資訊進行自動切分，得到文本對應的分詞序列，實現從文本到手語碼的轉換。手語運動資料庫記錄了每個手語詞的手語運動資訊，涵蓋了中國基本手語詞、同義手語詞、指拼手語詞等。手語碼經由手語運動資料庫的“翻譯”，合成手語動作。系統將手語動作以三維圖像的方式展示，且生成的手語圖像能夠被聽力語音障礙者便捷準確地閱讀和理解。

#### 3.2 語音識別

語音識別是指將人類的語音中的詞彙內容轉換為計算機可讀的輸入[6]。對於視頻講座的語音識別，主要是把講座的音頻轉化為文本，為下一步手語碼轉換提供信息。音頻的來源有兩種方式，見圖 3、圖 4。

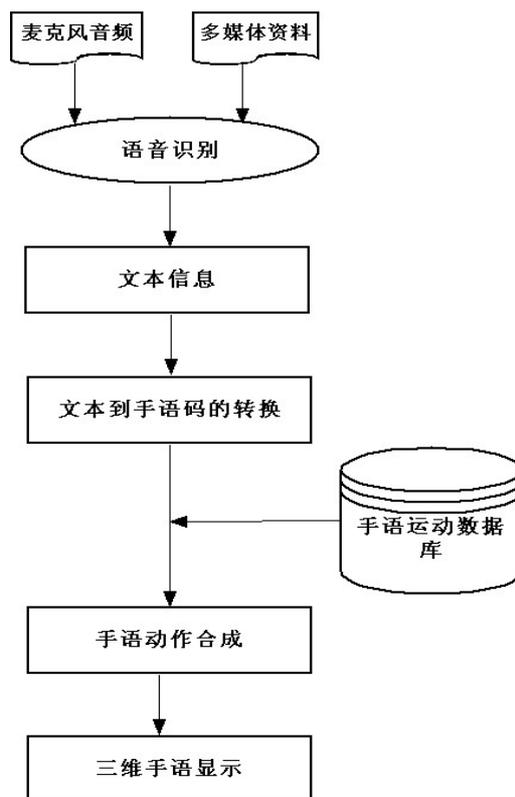


圖 2. 總體流程圖

Figure 2. Overall flow chart

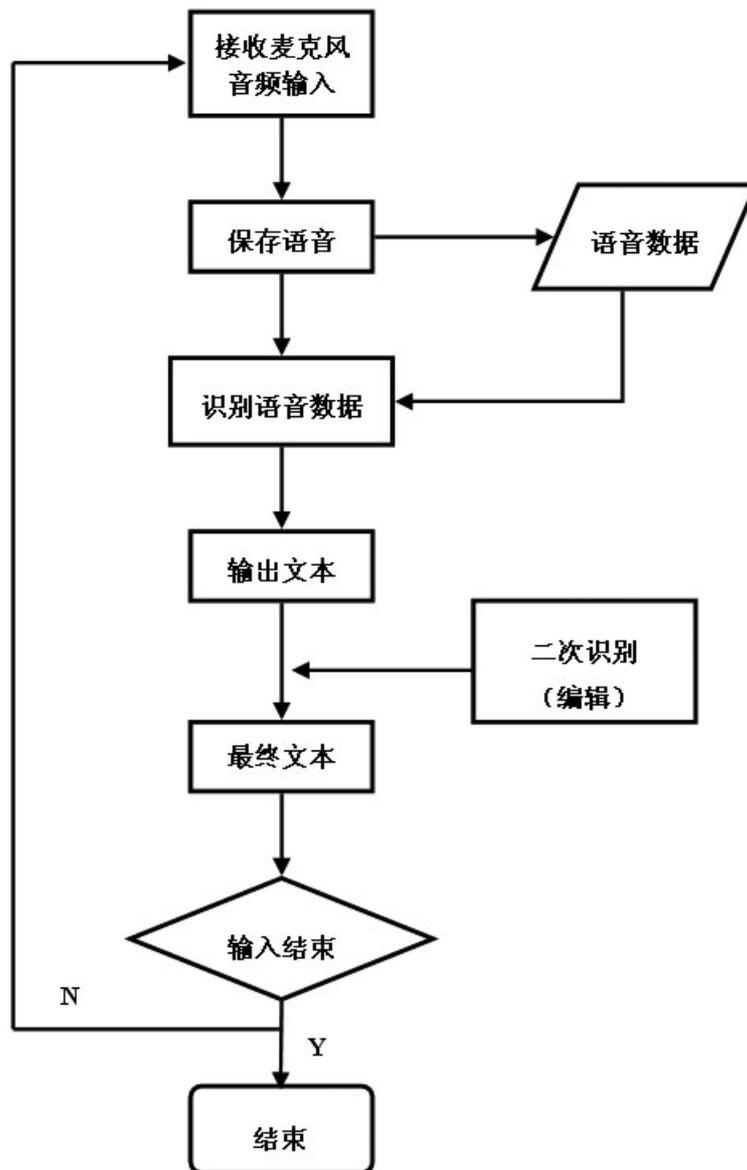


圖3. 即時語音識別的流程圖

Figure 3. A real time speech recognition flow chart

如圖 3 所示，說話人通過麥克風輸入語音，系統自動將當前聲音信號保存為語音數據，並對語音數據進行識別，實時輸出對應的文本。

語音識別是指將人類的語音中的辭彙內容轉換為電腦可讀的輸入[6]。對於視頻講座的語音識別，主要是把講座的音頻轉化為文本，為下一步手語碼轉換提供資訊。音頻的來源有兩種方式，見圖 3、圖 4。

如圖 3 所示，說話人通過麥克風輸入語音，系統自動將當前聲音信號保存為語音資料，並對語音資料進行識別，即時輸出對應的文本。

如圖4所示，系統輸入為多媒體資料檔，即視頻或音頻的編碼檔（視頻檔為Mpeg1/2/4和WMV格式，音頻為WAV/MP3格式），系統將檔解碼為原始語音資料，並將資料切分為若干小段，逐段進行語音識別，輸出對應的文本。

### 3.3 文本到手語碼的轉換

如圖5所示，文本句法分析到手語碼的轉換是手語合成中的關鍵環節，直接關係到手語內容的表意準確。中國手語是參考漢語制定的，但在基本詞彙、句子結

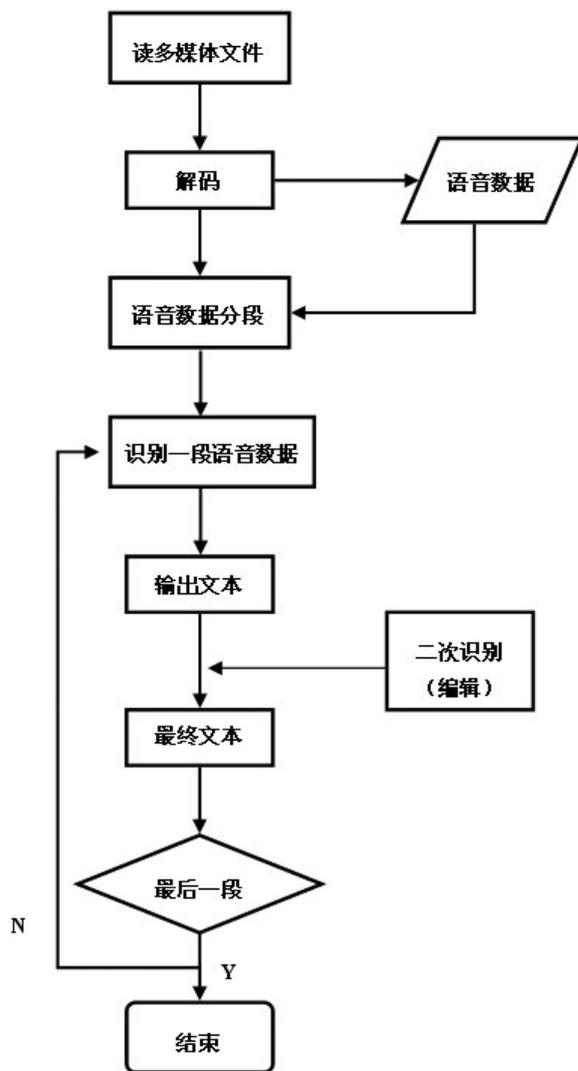


圖 4. 多媒體資料的語音識別流程圖  
Figure 4. A multimedia speech recognition flow chart

構和構詞方法上有著明顯的區別。中國手語的辭彙由 3330 個手勢語組成，遠少於漢語十幾萬的辭彙量，手勢詞語與漢語的詞語不完全存在一一對應的關係。在句子結構和構詞方法上，手語句子的語序以主謂賓為主[7]，在一些時候與漢語句子的詞序有所不同。手語合成中的文本分析過程是在語音識別的漢語語句的基礎上，按照手語的規則提取和組合用來表意手語碼的辭彙，組成手語碼序列。

### 3.4 手語運動資料庫的建立

手語運動資料庫的建立流程如圖 6。手語運動資料庫記錄了每個手語詞的手語運動資訊，是手語合成

的基礎。目前建立手語詞庫的方法有兩種：運動跟蹤方法[8][9]和手工編輯方法[10]。兩種方法各有利弊[3]，可以將這兩種方法結合，首先進行動作資料獲取，建立初始手語動作詞庫，然後對不準確的手語運動進行編輯修改。這樣可以獲得一個準確、逼真的手語詞庫，而且減少了工作量。

### 3.5 手語動作的合成

在經文本分析和手語運動資料檢索後，得到的手語符號序列中包含手語詞和手語運動參數，手語運動合成的目的是將這樣的符號序列轉換成手語動作。

如圖 7 所示，手語動作合成主要完成手語動作之

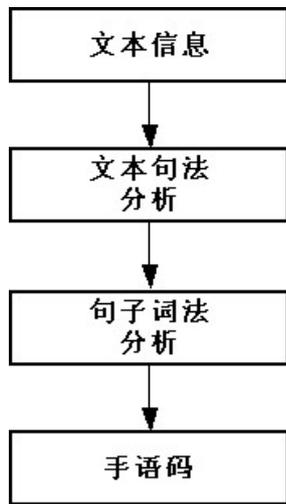


圖 5. 文本到手語碼轉換的流程圖

Figure 5. A text to sign language translation flow chart

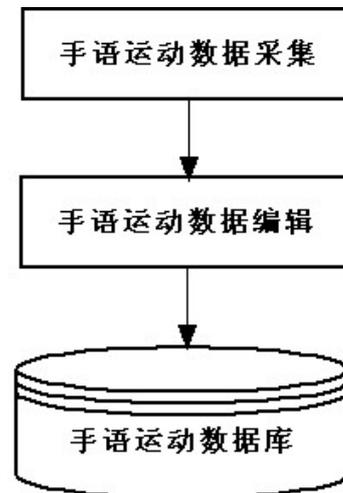


圖 6. 手語運動資料庫建立的流程圖

Figure 6. A hand signals database setup flow chart

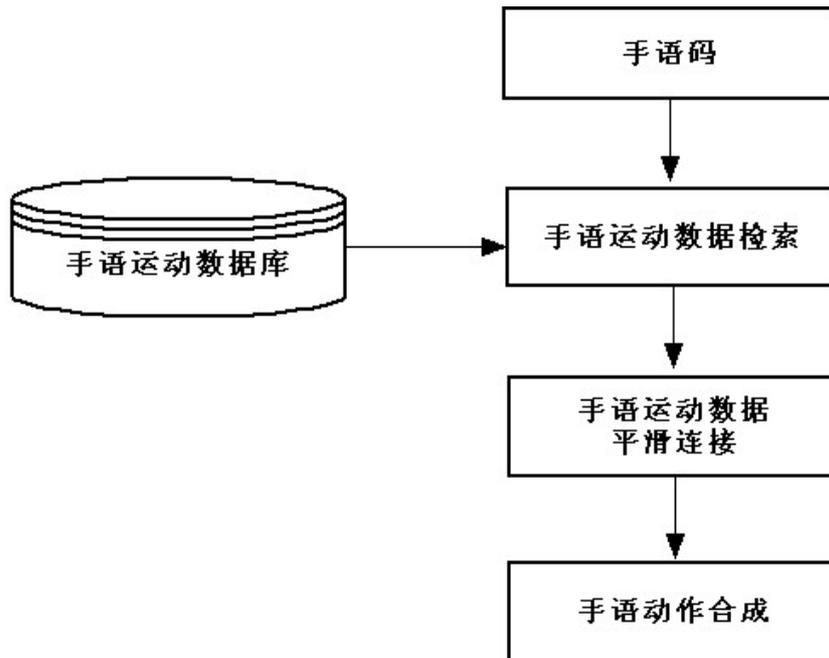


圖 7. 手語動作合成的流程圖

Figure 7. A hand signals synthesis flow chart

間的過渡和銜接，要求合成的手語運動序列具備連續性和平滑性，並要求不同的手語動作之間有較好的過渡動作。手語符號序列中的每個手語詞的手語運動資料可直接從手語詞運動資料庫中查詢得到，手語詞的錄製是完整且連貫的，因此其內部具有很好的運動逼真性和連續幀之間的運動相關性[3]。但，由於兩個連續的手語詞之間，前一個手語詞的最後一個手勢（最後一幀）與下一

個手語詞的第一個手勢（第一幀）之間的差別是任意的，因此手語詞之間的平滑過渡則要更複雜。

### 3.6 三維手語的顯示

手語是一種人之間交流的可視語言，手語動作最終以三維虛擬人的動作展示給語音障礙者。

如圖 8 所示，系統能夠以三維的方式展示手語，

且生成的手語圖像能夠被聽力和語音障礙者便捷準確地進行閱讀和理解。一個手語運動是一個人體手勢的序列，按照預定的時間間隔連續顯示一個手語運動中的每一個手勢，既可以生成對應的手語運動。手語三維顯示實際上就是根據給定的手語運動表示，驅動虛擬人的上肢運動，以此顯示逼真的手語[4]。

三維手語的顯示除了要滿足手語顯示的連貫性和準確性要求，還要考慮檔傳輸的需求。對於線上講座直播的三維手語展示，希望能夠在互聯網上線上使用，一般三維動畫技術形成的圖像由於檔體積大，不適合網路應用。這就對可以使用三維動畫技術來實現手語動畫，但一般的三維動畫技術形成的圖像雖然可能滿足生動逼真的要求，卻不適合網路應用。Virtual Reality Modeling Language (VRML)，是一種虛擬現實建模語言。它主要用於在 Internet 上表示並共用三維虛擬場景和虛擬物體[11]。在 VRML 中有一部分是專門用於描述三維人體模型 H-Anim 標準，根據此標準對虛擬人的定義，應用運動學的方法和電腦圖形學的方法就可以計算出虛擬人每個肢體的位置和方向，由此確定虛擬人的一個姿態，從而完成一系列動作[5]。

#### 4. 結束語

本文論述了一個正在研究中的國家數位圖書館的手語合成系統。該系統通過手語合成技術對國家圖書館的線上講座、視頻資料、現場講座等數位視頻資源進行

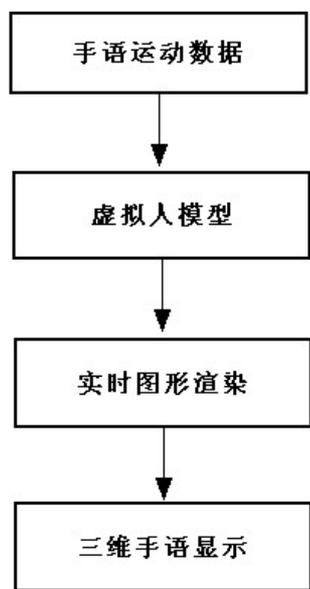


圖 8. 三維手語顯示的流程圖  
Figure 8. A flow chart of sign language 3D display

手語翻譯，並用三維虛擬人像手語展示，為廣大聾啞人做手語同步導視。這是手語系統的一個重要的應用領域和實踐，可以幫助有聽力和語音障礙的讀者和正常人一樣獲取國家數字圖書館的資源。

#### 5. 附錄：術語

**聽覺和語音障礙者[1]**：指那些由於疾病、傷害等各種原因導致聽力障礙和發音障礙的人，從而不能同一般人進行正常的話語交往活動，也即我們通常說的“聾啞人”。

**語音識別**：就是根據輸入的語音資料，系統自動地把語音翻譯成對應的文本。

**手語合成**：就是系統根據輸入的文本資訊，自動把文本資訊翻譯成對應的手語資訊，並採用圖形圖像技術，由電腦生成的虛擬人把手語逼真地展示出來。

#### REFERENCES

- [1] Fu Yiting, Mei Cikai (1986). An Outline of Sign Language of the Deaf. (in Chinese) (付逸亭, 梅次開. 聾人手語概論. 學林出版社, 1986.)
- [2] China National Association of the Deaf (1988). Chinese Sign Language. (in Chinese) (中國聾人協會. 中國手語. 華夏出版社, 1988.)
- [3] Wang Zhaoqi, Gao Wenji (2002). The Synthesis Approches of Chinese Sign Language Based on Dummy Man Synthesis Technology, Journal of Software, 13(10), 2051-2056 (in Chinese) (王兆其, 高文基. 於虛擬人合成技術的中國手語合成方法. 軟件學報, 2002, 13(10), 2051-2056.)
- [4] Wang Zhaoqi, Yang Changshui, Gao Wen (2003). 3D Display of Chinese Sign Language Based on VRML. Computer Study and Development. 40(3), 459-463. (in Chinese) (王兆其, 楊長水, 高文. 基於 VRML 的中國手語三維顯示. 電腦研究與發展. 3/2003, 40(3), 459-463.)
- [5] Gao Wei, Guo Jin, Zeng Peikai (2002). The Current Situation and Future Consideration of the Sign Language Research. Journal of IT Technology, 6-8. (in Chinese) (高偉, 郭瑾, 曾碩凱. 手語研究的方向與現狀. 電子技術應用, 11/2002, 6-8.)
- [6] Wikipedia Editor : Speech Recognition 2009/06/02 23:08[2009/07/07.06:55]  
[http://zh.wikipedia.org/w/index.php?title=E8%AF%AD%E9%9F%B3%E8%AF%86%E5%88%AB&oldid=10222296.](http://zh.wikipedia.org/w/index.php?title=E8%AF%AD%E9%9F%B3%E8%AF%86%E5%88%AB&oldid=10222296) (in Chinese) (維琪百科編者. 語音識別[G/OL]. 維琪百科, 自由的百科全書, 2009年06月2日23:08 [2009年07月7日06:55]. [http://zh.wikipedia.org/w/index.php?title=E8%AF%AD%E9%9F%B3%E8%AF%86%E5%88%AB&oldid=10222296.](http://zh.wikipedia.org/w/index.php?title=E8%AF%AD%E9%9F%B3%E8%AF%86%E5%88%AB&oldid=10222296))
- [7] Wikipedia Editor : Sign Language 2009/06/13 06:36[2009/07/02 06:32]  
<http://zh.wikipedia.org/w/index.php?title=Sign%20Language>

- title=%E6%89%8B%E8%AA%9E&oldid=10305654.) (in Chinese) (維琪百科編者. 手語[G/OL]. 維琪百科, 自由的百科全書, 2009年06月13日06:36 [2009年07月2日06:32] <http://zh.wikipedia.org/w/index.php?title=%E6%89%8B%E8%AA%9E&oldid=10305654>.)
- [8] Erenshteyn, R., Laskov, P., Foulds, R., *et al.* Recognition approach to gesture language understanding. In: Kavanagh, M. E., Werner, B., eds. Proceedings of the 13th International Conference on Pattern Recognition. Los Alamitos: IEEE Computer Society, 1996, 431-435.
- [9] Fels, S., Hinton, G. Glove-TalkII: A neural network interface which maps gestures to parallel formant speech synthesizer controls. *IEEE Transactions on Neural Networks*, 1998, 9(1): 205-212.
- [10] Noriko, T., *et al.* An alternative method for building a database for American sign language. In: Proceedings of the Technology for Persons with Disabilities Conference 2000. Los Angeles, CA, 2000. <http://www.csun.edu/cod/conf/2000/proceedings/0071Tomuro.htm>.
- [11] VRML97 Specification ISO/IEC DIS 14772.