

语音合成

維基百科，自由的百科全书

語音合成是將人類語音用人工的方式所產生。若是將電腦系統用在語音合成上，則稱為語音合成器，而語音合成器可以用軟/硬體所實現。文字轉語音（Text-To-Speech, TTS）系統則是將一般語言的文字轉換為語音，其他的系統可以描繪語言符號的表示方式，就像音標轉換至語音一樣^[1]。

而合成後的語音則是利用在資料庫內的許多已錄好的語音連接起來。系統則因為儲存的語音單元大小不同而有所差異，若是要儲存phone以及diphone的話，系統必須提供大量的儲存空間，但是在語意上或許會不清楚。而用在特定的使用領域上，儲存整字或整句的方式可以達到高品質的語音輸出。另外，包含了聲道模型以及其他的人類聲音特徵參數的合成器則可以創造出完整的合成聲音輸出^[2]。

一個語音合成器的品質通常是決定於人聲的相似度以及語意是否能被了解。一個清晰的文字轉語音程式應該提供人類在視覺受到傷害或是得到失讀症時，能夠聽到並且在個人電腦上完成工作。從80年代早期開始，許多的電腦作業系統已經包含了語音合成器了。

目录

文字處理的概觀

歷史

合成器技術

串接合成

單位選擇合成

Diphone合成

Domain-specific合成

共振峰合成

發音合成

HMM基礎合成

正弦波合成

深度學習合成

開源代码

開源数据

英文

中文

粵語

經典論文

相關比賽

含有語音合成器的軟體

語音合成標記語言

參見

參考文獻

外部連結

語音合成服務提供商

文字處理的概觀

歷史

早在17世紀就有法國人研發機械式的說話裝置。直到19世紀，貝爾實驗室對於電子語音合成技術的研究，才開啟近代語音合成技術的發展。貝爾實驗室在1939年製作出第一個電子語音合成器 VODER^[3]，是一種利用共振峰原理所製作的合成器。1960年，瑞典語言學家G. Fant則提出利用線性預測編碼技術（LPC）來作為語音合成分析技術，並推動了日後的發展。後來1980年代Moulines E和Charpentier F提出新的語音合成演算法PSOLA，此技術可以合成比較自然的語音。

合成器技術

串接合成

單位選擇合成

這通常是使用字詞預錄的方式，將各字詞的發音預先錄製好，經處理後存放在資料庫中。使用時再將預發音的字句斷成字詞，再把發音的資料自資料庫中取出，交由DSP去處理發音的合成。

Diphone合成

Domain-specific合成

共振峰合成

發音合成

HMM基礎合成

正弦波合成

深度學習合成

深度學習語音合成使用深度神經網絡（DNN）從文本（文本到語音）或頻譜（聲碼器）產生人工語音。深度神經網絡使用大量錄製的語音進行訓練，如果是文本到語音系統，則使用和語音相對的文本。

基於DNN的語音合成器正在接近人聲的自然度。該方法的缺點的例子是當數據不充分時強健性較低，缺乏可控性，自動回歸模型的性能低。一些局限性（如缺乏可控性）可以通過未來的研究得到解決。

目前以 Tacotron2 (https://pytorch.org/hub/nvidia_deeplearningexamples_tacotron2/) + Waveglow (<https://arxiv.org/abs/1811.00002>) 只需要數十個小時的錄音語料加以訓練即可合成出品質非常高的語音。但是對於聲調語言(如中文或台語)，因為有不同程度的變調需求，有時合成的語音會出現變調錯誤的情況。

开源代码

参见 Github 语音合成主题^[4]

开源代码

类别	语言	功能	名称	网址
前端 Frontend	英文	文本正则化Text Normalization		
		转拼音 G2P(Grapheme to Phoneme)		
		韵律预测Prosody Prediction	Predicting Prosodic Prominence from Text	https://github.com/Helsinki-NLP/prosody
	Find prosodic events from wav		https://github.com/asuni/wavelet_prosody_toolkit	
中文		g2pC	https://github.com/Kyubyong/g2pC	
声码器 Vocoder	/			

开源数据

英文

*LJSpeech^[5]

中文

*标贝科技开源的中文女声语音合成数据^[6]

*爱数智慧开源的15分钟中文童声数据^[7]

粤语

暂无

经典论文

Github论文收集库：[awesome-speech-recognition-speech-synthesis-papers^{\[8\]}](#)

相关比赛

- [Blizzard Challenge^{\[9\]\[10\]}](#)

- [voice conversion](#)^[11]

含有語音合成器的軟體

- [NVDA](#)（[开源多国语言螢幕報讀軟體](#)）
- [台灣工研院資通所語音團隊開發出的mTTS](#)
- [JAWS](#)（[商業付費螢幕報讀軟體](#)）
- 內建於蘋果電腦相關產品的[VoiceOver Utility](#)
- [Android 1.6版](#)加入支援語音合成的功能。^[12]
- 微軟在1995年開始，開發了一套[Microsoft Speech API](#)，微軟的一些作業系統或軟體有提供支援語音功能。
- [VOCALOID](#) 及 [VOICEROID](#)，前者偏重歌唱而後者偏重 TTS。
- [盛大遊戲](#)在2012年初，發佈了Android語音合成平臺聽聽中心，支援本地中英文合成及變聲等服務。
- [台語語音筆記本](#) (<https://web.archive.org/web/20180315200332/http://vikon.myweb.hinet.net/mltmenu3.htm>)^[13]為台灣話的語音輸出 (Taiwanese Text-to-speech) 軟體。系統使用[台語聲調群剖析器](#) (<https://web.archive.org/web/20180704092844/http://vikon.myweb.hinet.net/ttgpc.htm>) (Taiwanese tone group parser)^[14]、[語音引擎](#) (Speech engine) 及 [語音合成器](#) (Speech synthesizer)直接安裝於個人電腦獨立運作，不需連結 MS Speech SDK或 IBM TTS Engine。使用者圖行形介包含可用台語羅馬字或中文輸入、有聲辭典同步顯示、有聲讀本製作、外部應用程式/網路瀏覽器語音輸出、台語調符轉換及使用中英文語詞搜尋等功能。

語音合成標記語言

参见

- [人机界面](#)
- [語音识别](#)
- [手写识别](#)
- [VOCALOID](#)
- [初音未來](#)
- [智能儀錶](#)
- [智能玩具](#)
- [电子地图](#)
- [电子导游](#)
- [电子词典](#)

参考文献

1. Allen, Jonathan; Hunnicutt, M. Sharon; Klatt, Dennis. From Text to Speech: The MITalk system. Cambridge University Press. 1987. ISBN 978-0-521-30641-6.
2. Rubin, P.; Baer, T.; Mermelstein, P. An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America*. 1981, **70** (2): 321–328. Bibcode:1981ASAJ...70..321R. doi:10.1121/1.386780.
3. 柳春、于洪志（2008年）：《現代教育技術》，第26期第64頁，北京：清華大學電教中心。
4. github-tts-topic. （原始内容存档于2020-05-30） .
5. LJ-Speech-Dataset. （原始内容存档于2020-06-04） .
6. 数据描述. （原始内容存档于2020-05-02） .
7. 数据页面. [2020-05-30].
8. ponyzhang, zzw922cn/awesome-speech-recognition-speech-synthesis-papers, 2020-05-29 [2020-05-30], （原始内容存档于2020-07-09）
9. https://www.synsig.org/index.php/Blizzard_Challenge_2020
10. 存档副本. [2020-05-28]. （原始内容存档于2020-04-23） .
11. 存档副本. [2020-05-28]. （原始内容存档于2020-02-19） .
12. Jean-Michel Trivi. An introduction to Text-To-Speech in Android. *Android-developers.blogspot.com*. 2009-09-23 [2010-02-17]. （原始内容存档于2011-07-08） .
13. 田村志津枝 (https://ja.wikipedia.org/wiki/%E7%94%B0%E6%9D%91%E5%BF%97%E6%B4%A5%E6%9E%9D) （页面存档备份 (https://web.archive.org/web/20181011093411/https://ja.wikipedia.org/wiki/%E7%94%B0%E6%9D%91%E5%BF%97%E6%B4%A5%E6%9E%9D)，存于互联网档案馆）（2010）。初めて台湾語をパソコンに喋らせた男—母語を蘇らせる物語。東京：現代書館。[Tamura, S.(2010).Hajimete Taiwango o pasokon ni shaberaseta otoko: bogo o yomigaeraseru monogatari.Tokyo, Japan: Gendai Shokan.]
14. Chang, Y. C. (2017). 以知識表徵方法建構台語聲調群剖析器 (A Knowledge Representation Method to Implement A Taiwanese Tone Group Parser)[In Chinese]. *International Journal of Computational Linguistics & Chinese Language Processing*; 22:2 2017.12[民106.12]; 頁73-86

外部連結

- TTS - ProgWiki (http://www.player.idv.tw/prog/index.php?title=TTS) （页面存档备份 (https://web.archive.org/web/20080926153555/http://www.player.idv.tw/prog/index.php?title=TTS)，存于互联网档案馆）
- Ekho（余音）中文开源语音合成软件 (http://www.eguidedog.net/ekho.php) （页面存档备份 (https://web.archive.org/web/20100907185749/http://www.eguidedog.net/ekho.php)，存于互联网档案馆）
- 工研院語音合成器 (https://web.archive.org/web/20110624151743/http://atc.ccl.itri.org.tw/)
- 工研院文字轉語音Web服務 (https://web.archive.org/web/20120310172450/http://tts.itri.org.tw/)
- 盛大語音合成平臺 (https://web.archive.org/web/20120226083611/http://tingting.sdo.com/)
- 台語語音筆記本 (https://web.archive.org/web/20180315200332/http://vikon.myweb.hinet.net/mltmenu3.htm)

语音合成服务提供商

- [科大讯飞 语音合成体验网站 \(https://www.xfyun.cn/services/online_tts\)](https://www.xfyun.cn/services/online_tts) (页面存档备份 (https://web.archive.org/web/20200718102835/https://www.xfyun.cn/services/online_tts), 存于互联网档案馆)
- [阿里巴巴 语音合成体验网站 \(https://ai.aliyun.com/nls/tts\)](https://ai.aliyun.com/nls/tts) (页面存档备份 (<https://web.archive.org/web/20200514041254/https://ai.aliyun.com/nls/tts>), 存于互联网档案馆)
- [百度 语音合成体验网站 \(https://ai.baidu.com/tech/speech/tts\)](https://ai.baidu.com/tech/speech/tts) (页面存档备份 (<https://web.archive.org/web/20200605225826/https://ai.baidu.com/tech/speech/tts>), 存于互联网档案馆)
- [腾讯 语音合成体验网站 \(https://cloud.tencent.com/product/tts\)](https://cloud.tencent.com/product/tts) (页面存档备份 (<https://web.archive.org/web/20200617045555/https://cloud.tencent.com/product/tts>), 存于互联网档案馆)
- [微软 语音合成体验网站 \(https://azure.microsoft.com/zh-cn/services/cognitive-services/text-to-speech/#features\)](https://azure.microsoft.com/zh-cn/services/cognitive-services/text-to-speech/#features) (页面存档备份 (<https://web.archive.org/web/20200721002411/https://azure.microsoft.com/zh-cn/services/cognitive-services/text-to-speech/#features>), 存于互联网档案馆)
- [标贝科技 语音合成体验网站 \(https://www.data-baker.com/tts_en.html\)](https://www.data-baker.com/tts_en.html) (页面存档备份 (https://web.archive.org/web/20180821101446/https://www.data-baker.com/tts_en.html), 存于互联网档案馆)
- [京东 语音合成体验网站 \(https://neuhub.jd.com/ai/api/speech/tts\)](https://neuhub.jd.com/ai/api/speech/tts) (页面存档备份 (<https://web.archive.org/web/20200114095121/https://neuhub.jd.com/ai/api/speech/tts>), 存于互联网档案馆)
- [极限元 语音合成体验网站 \(http://experience.jixianyuan.com/tts/tts.html\)](http://experience.jixianyuan.com/tts/tts.html) (页面存档备份 (<https://web.archive.org/web/20200221173018/http://experience.jixianyuan.com/tts/tts.html>), 存于互联网档案馆)
- [Rokid 语音合成体验网站 \(https://developer.rokid.com/#/tts\)](https://developer.rokid.com/#/tts) (页面存档备份 (<https://web.archive.org/web/20200408072043/https://developer.rokid.com/#/tts>), 存于互联网档案馆)

取自“<https://zh.wikipedia.org/w/index.php?title=语音合成&oldid=72051835>”

本页面最后修订于2022年6月8日 (星期三) 03:08。

本站的全部文字在知识共享 署名-相同方式共享 3.0协议之条款下提供，附加条款亦可能应用。（请参阅使用条款）
Wikipedia®和维基百科标志是维基媒体基金会的注册商标；维基™是维基媒体基金会的商标。
维基媒体基金会是按美国国内稅收法501(c)(3)登记的非营利慈善机构。