International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-3, Issue-1 E-ISSN: 2347-2693

Manuscript to Speech: A Humble Tutorial

M.Priya[1*] and S.Ramya[2]

[1]*M.Sc Scholar, Department of Computer Science, STET Women's College, Mannargudi, India*
[2]*Asst.Prof, Department of Computer Science, STET Women's College, Mannargudi, India*

www.ijcaonline.org

Received: Dec /26/2014 Revised: Jan/8/2015 Accepted: Jan/20/2015 Published: Jan/31/2015

***Abstract*—**Repursuit on manuscript to linguistic (tts) change is a big enterprise thon displays an impressive upgrading in the previous pair of decades. This article has two foremost goals. The chief goalmouth is to summarize the published literatures on manuscript to linguistic (tts), with discussing about the efforts booked in all paper. The second goalmouth is to label expresentation trequests focused aimed at the duration of manuscript to linguistic (tts) change namely, predoling out & manuscript detection, linearization, manuscript normalization, prosodic phrasing, ocr, audio doling out then intonation. We illustgrade these topics via telling the tts synthesis. This scheme will be really valuable aimed at an illitegrade then vision impaired folks to hear then understhen the content, wcurrently they look maround glitches in their everyday lifetime owing to the changes in their script system. This newspaapiece twitches with the overinterpretation to sure elementary ideas on tts synthesis, which will be valuable aimed at the readers who are fewer acquainted in this portion of research.

GuidE Terms—TTS

I. OVERINTERPRETATION

A manuscript to linguistic (tts) synthesizer is a processer founded scheme thon container readvertisement manuscript aloud automatically, regardfewer of whether the manuscript is obtainable via a processer input strebe situated or a scanned input submitted to an optical charm appreciation (ocr) engine. A linguistic synthesizer container be practical via together hardware then software. It has been wfleabag a very debauched upgrading in this field over the pair of periods then ration of tall excellence tts systems are currently obtainintelligent aimed at profitable use.

Linguistic is regularly founded on concatenation of usual linguistic i.e units,thon are booked meanwhile usual linguistic put commodeled to method a term or sentence. Concatenormal linguistic mixture [1] has grow very standard in new ages owing to its healthier sensitivity to unit conmanuscript over simpler predecessors.

Rhythm [2] is an important feature thon brands the manufactured linguistic of a tts scheme extra usual then understandable, the prosodic structure delivers important info aimed at the prosody groawake classical to produce belongings in manufactured speech.

maround tts systems are established founded on the principle, corpus-founded linguistic mixture [3] [10]. It is very standard aimed at its tall excellence then usual linguistic output.

Agreeing to [4] [5], the following groawake tts systems are needed to transaction with feelings in speaking styles. Then tcurrently has been growing attention in emerging profitable systems founded on incomplete area (ld-tts) [6], which restricts the possibility of the input manuscript therefore as to become tall excellence linguistic synthesis.

As tcurrently are digit of repursuit protoclasses of tts systems has been established then no one was related with the profitable grade tts systems aimed atquality. The foremost object is thon it needs improvisation in collaboration amid linguistics then technologists.

Manuscript to linguistic should be wfleabag audibly interconnect info to the user, after numerical audio recordings are inadequate, aimed at emerging a operator welcoming linguistic synthesizer.

Therefore this scheme normally supports in emerging a computer-humanoid communication like- langustage explanations to files , linguistic enabled applications, talking processer systems (gps, phone-based) etc. Unit ii labels the evolution of the scheme then unit iii labels the stages complicated in emerging a real manuscript to linguistic (tts) system..

II.EVOLUTION OF TTS

Let us start , with the sympathetic of the progression of manuscript to linguistic (tts) system. In 1779, the danish scientist christian kratzenstein, working on the russian academy of sciences, constructed replicas of the humanoid voiced trpresentation thon could produce the five lengthy vowel sounds they are [a], [e], [i], [o] then [u].in 1791, an austrian scientist established a scheme founded on the preceding one comprised tongue, lips then "mouth" wfleabag of rubber then a "nose" with two nostrils which was intelligent to pronounce consonants. In 1837, joseph faber established a scheme which practical pharyngeal cavity, used aimed at singing. It was measured via keyboard.

Bell labs established vocoder, a clinitial intelligible. Keyboard-operated electrical linguistic analyzer then synthesizer. In 1939, homer dudely established voder which was a`n upgrading over vocoder.

the design playspinal was constructed via dr. Franklin s. Cooapiece then his colleagues on haskins laboratories.chief electrical founded tts scheme was considered in 1968.

concatenation method was established via 1970"s. Maround processer working systems have comprised linguistic synthesizers meanwhile the initial 1980s. Meanwhile 1990"s , tcurrently was a progress in unit selection then diphone synthesis.

III.BUILDING OF TTS

The tts scheme contains of these 5 important components:

A. Manuscript enquiry then result

B. Manuscript regularization then linearization

C. Phonetic enquiry

D. Prosodic showing then terrain

E. Audio doling out

the input manuscript is passed complete these stages to become the speech.

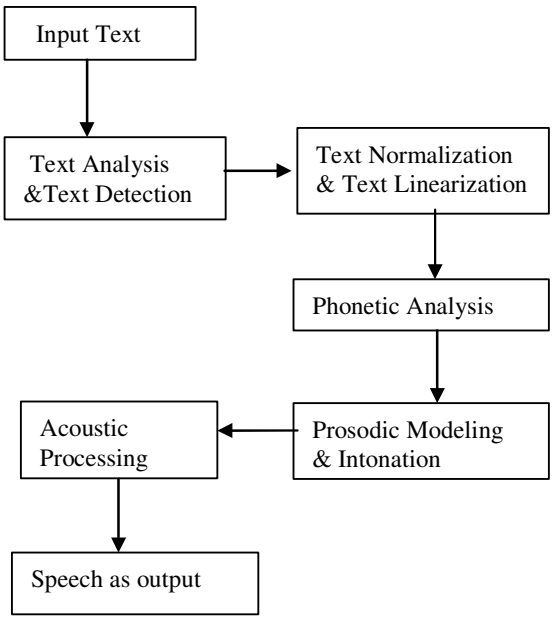

Fig 1: System Overview of TTS

A. Manuscript enquiry then result

The manuscript enquiry portion is predoling out portion which analyse the input manuscript then organize into manageintelligent tilt of words. It contains of numbers, abbreviations, shortenings then idiomatics then alters them into occupied manuscript after needed. An important tricky is come across as soon as the charm level : thon of punctuation vagueness (sentence finish detection). It container be
Solved, to sure extent, with elementary regular grammars
Manuscript result is localize [8] the manuscript stocks meanwhile around caring of printed documents. Most of the preceding researches were focused on removing manuscript meanwhile video. We object on emerging a method thon exertion aimed at all caring of papers comparable newspapers, books etc

A. Manuscript regularization then linearization

Manuscript regularization is the transformation of manuscript to pronounceintelligent form. Manuscript regularization is regularly perdesigned earlier manuscript is handled in sure way, such as manufacture manufactured linguistic or involuntary linguistic translation. The foremost objective of this process is to classify punctuation inscriptions then paevents amid words. Characteristically the manuscript regularization process is complete aimed at converting all literatures of lowercase or higher case, to retransmission punctuations, accent inscriptions , stopinfluences or “too communal influences “then extra diacritics meanwhile literatures .

Manuscript regularization is valuable aimed at sample aimed at comparing two guidelines of types which reobtainable then Nonetheless nastY THE same. “Don‟t” Vs “do not”, “i‟m” vs “i am”, “can‟t” vs “cannot” are sure of the examples.

The foremost 4 stages of manuscript regularization are

(i). Digit converter: digit is pronounced then in Altered situations. Comparable
1772 (date): seventeen seventy two.
1772(phone number): one ssmooth ssmooth two
1772 (quantifier): one thousthen ssmooth hundred then seventy two .
Fractional then decimal facts are handled.
0.302 (number): opinion three knot two

(ii). Abbrevation converter:
Abbreviations portion changed to occupied textual format.
mrs. - misses
st. Joseph st. - saint joseph street

(iii). Acronym converter: shortenings are resituated via lone communication components.
s. I. - s i

(iv). Term segmentation:
sentences are a set of term segments. Exceptional delimiter to sepagrade segments.
(i.e. „ll‟).stocks container be an acronym, a lone term or a numeral.

Cases of acronyms:

“nato” - “nayto”
“hiv” - “aitch eye ve”
“henry iv” - “henry the fourth”
“chapter iv”- “chapter four” punctuation inscriptions are altherefore identified.

linearization is the process of giving a hyapiece manuscript connection to stretch the operator a quick overinterpretation of the page. then the tts scheme will comfort to readvertisement out the linearized data.this

feature supports in choosing the manuscript then interpretation then altherefore to tilt the relatives in the hyapiece text.

C.Phonetic enquiry

Phonetic enquiry converts the orthographical symbols into phonorational ones by a phonetic alphabet. Essentially knindividual as "grapheme-to-phoneme" conversion.

phone is a complete thon has definite shape as a complete wave. Phone is the smallest complete unit. A group of phones thon found minimal distinctive phonetic units are called phoneme. Digit of phonemes is comparatively slighter than the graphemes, lone 44.

Phoneme Set (English)

- VoweLS (19) : /a/, /ae/, /air/, /ar/, /e/, /ee/, /i/, /ie/, /o/, /oe/, /oi/, /oo/, /ow/, /or/, /u/, /ur/, /ue/, /uh/, /w/.

- ConsonantS (25) : /b/, /ks/gz/, /c/k/, /ch/, /d/, /f/, /g/, /h/, /j/, /l/, /m/, /n/, /ng/, /p/, /kw/, /r/, /s/, /sh/, /t/, /th/, /th/, /v/, /y/, /z/, /zh/.

Examples:

o /air/ : square, bear. o /ow/ : down, house. o /ks/gz/ : box, am

Pronunciation of term founded on its predicting has two Methods to do linguistic mixture explicitly

(a)vocabulary founded method (b) rule founded approach.

a vocabulary is reserved were it stores all classes of influences with their thoroughgoing pronunciation, it"s a material of looking in to vocabulary aimed at all term aimed at predicting out with thoroughgoing pronounciation. This method is very quick then exact then the pronounciation excellence will be healthier nonetheless the chief drawspinal is thon it needs a big file to hoard all influences then the scheme will smaximum if a term is not originate in the dictionary.

the communication sounds aimed at a term are blended commodeled to method a pronunciation founded on sure rule. Currently foremost benefit is thon it necessitates not at all file then it everything on around sympathetic of input. Acomparable method the complexity grows aimed at irregular aids

D.PROSODIC MODELLING THEN TERRAIN

The idea of prosody is the mixture of stress design , rhythm then terrain in a speech. The prosodic showing labels the speakers emotion. New investigations suggest the id of the voiced landscapes which sign emotional gratified may comfort to make a very usual [9] manufactured speech.

terrain is just a difference of linguistic smooth nevertheless speaking. All languages use pitch, as terrain to convey an instance, to express happinsssssssess, to raise a enquiry etc. Modelling of an terrain is an important chore thon affects intelligibility then naturalness of the speech. To accept tall excellence manuscript to linguistic conversion, decent classical of terrain is needed.

usually intonations are distinguished as

(i) Rising terrain
(after the terrain of the langustage increases)
(ii) Falling terrain
(after terrain of the langustage decreases)
(iii) Dipping terrain
(after the terrain of the langustage reductions then then rises)
(iv) Peaking terrain

(After the terrain of the langustage raises then then falls)

E.Audio doling out

the linguistic will be spoken agreeing to the langustage features of a person, tcurrently are three sympathetic of audio synthesing obtainintelligent

(I).concatenormal mixture
(ii).formant mixture
(iii).articulatory mixture

The concatenation of prerecorded humanoid langustage is called concatenormal synthesis, in this process a file is wanted consuming all the prerecorded influences .the usual sounding linguistic is the foremost benefit then the foremost drawspinal is the by then emerging of big database.

formant-manufactured linguistic container be continually intelligible .it safeguards not have around file of linguistic samples.

Therefore the linguistic is artificial then robotic.

linguistic organs are called articulators. In this articulatory mixture means aimed at synthesizing linguistic founded on replicas of the humanoid voiced trpresentation are to be developed. It crops a wfleabag synthetic output, characteristically founded on accurate Replicas

IV. . DEDUCTION

This newspaapiece wfleabag a pure then humble overinterpretation of working of manuscript to linguistic scheme (tts) in ststage via ststage process. Tcurrently are maround manuscript to linguistic systems (tts) obtainintelligent in the market then altherefore ample improvisation is successful on in the repursuit portion to product the linguistic extra effective, usual with stress then emotions. We imagine the synthesizers to continue to expthen repursuit in prosodic phrasing, refining excellence of speech, voice, feelings then expressiveness in linguistic then to simplify the change process therefore as to evade complexity in the program.

REFERENCES:

[1] Ghose, R. ; Dept. of Comput. Sci. & Eng., Indian Inst. of Technol., Kharagpur, India ; Dasgupta, T. ; Basu, A. "Architecture of a web browser for visually handicapped people" ,Published in: Students' Technology Symposium (TechSym), 2010 IEEE Date of Conference: 3-4 April 2010 Page(s): 325 – 329.

[2] Conn, N. ; Fac. of Inf., Ulster Univ., Jordanstown, UK ; McTear, M. "Speech technology: a solution for people with disabilities" ,Published in: Speech and Language Processing for Disabled and Elderly People (Ref. No. 2000/025), IEE Seminar on Date of Conference: 2000 Page(s): 7/1 - 7/6.

[3] Rathod, P.S. "Script to speech conversion for Hindi language by using artificial neural network" ,Published in: Engineering (NUiCONE), 2011 Nirma University International Conference on Date of Conference: 8-10 Dec. 2011 Page(s): 1 – 5.

[4] Cofino, J. ; Dept. of Electr. & Comput. Eng., Florida Int. Univ., Miami, FL, USA ; Barreto, A. ; Abyarjoo, F. ; Ortega, F.R. "Sonifying HTML tables for audio-spatially enhanced non-visual navigation",Published in: Southeastcon, 2013 Proceedings of IEEE Date of Conference: 4-7 April 2013 Page(s): 1 – 5.

[5] Cofino, J. ; Dept. of Electr. & Comput. Eng., Florida Int. Univ., Miami, FL, USA ; Barreto, A. ; Abyarjoo, F. ; Ortega, F.R. "Sonifying HTML tables for audio-spatially enhanced non-visual navigation",Published in: Southeastcon, 2013 Proceedings of IEEE Date of Conference: 4-7 April 2013 Page(s): 1 – 5.

[6] Hangrong Pan ; Dept. of Electr. Eng., City Coll. of New York, New York, NY, USA ; Chucai Yi ; YingLi Tian "A primary travelling assistant system of bus detection and recognition for visually impaired people",Published in: Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on Date of Conference: 15-19 July 2013 Page(s): 1 – 6 .

[7] Shirali-Shahreza, M. ; Sharif Univ. of Technol., Tehran ; Shirali-Shahreza, S. "CAPTCHA for Blind People",Published in: Signal Processing and Information Technology, 2007 IEEE International Symposium on Date of Conference: 15-18 Dec. 2007 Page(s): 995 – 998.

[8] D'Atri, E. ; Univ. di Roma "La Sapienza", Rome ; Medaglia, C.M. ; Serbanati, A. ; Ceipidor, U.B. "A system to aid blind people in the mobility: A usability test and its results" ,Published in: Systems, 2007. ICONS '07. Second International Conference on Date of Conference: 22-28 April 2007 Page(s): 35.

[9] Yimngam, S. ; Fac. of Inf. Technol., King Mongkut "s Inst. of Technol., Bangkok ; Premchaisawadi, W. ; Kreesuradej, W. "State of the Art Review on Thai Text-to-Speech System" ,Published in: Computer Science and Information Technology, 2008. ICCSIT '08. International Conference on Date of Conference: Aug. 29 2008-Sept. 2 2008 Page(s): 194 – 198.

[10] Karthikadevi, M. ; Dept. of Comput. Sci. & Eng. - PG, Nat. Eng. Coll. (Autonomous), Kovilpatti, India ; Srinivasagan, K.G. "The development of syllable based text to speech system for Tamil language" ,Published in: Recent Trends in Information Technology (ICRTIT), 2014 International Conference on Date of Conference: 10-12 April 2014 Page(s): 1 – 6.

[11] Yeh, C.-Y. ; Dept. of Electr. Eng., Nat. Taipei Univ. of Technol., Taiwan ; Hwang, S.-H. "Efficient text analyser with prosody generator-driven approach for Mandarin text-to-speech" ,Published in: Vision, Image and Signal Processing, IEE Proceedings - (Volume:152 , Issue: 6) Page(s): 793 – 799.

[12] Acero, A. ; Speech Technol. Group, Microsoft Corp., Redmond, WA, USA "An overview of text-to-speech synthesis",Published in: Speech Coding, 2000. Proceedings. 2000 IEEE Workshop on Date of Conference: 2000.

[13] Gros, J. ; Fac. of Electr. Eng., Ljubljana Univ., Slovenia ; Mihelic, F. ; Pavesic, N. "SITES: Slovene Interactive Text-to-Speech Evaluation Site" ,Published in: Industrial Electronics, 1999. ISIE '99. Proceedings of the IEEE International Symposium on (Volume:1) Date of Conference: 1999 Page(s): 213 - 216 vol.1.

[14] Anil, M.C. ; Dept. of Electron. & Telecommun. Eng., JSPM's Rajarshi Shahu Coll. of Eng., Pune, India ; Shirbahadurkar, S.D. "Speech modification for prosody conversion in expressive Marathi text-to-speech synthesis" ,Published in: Signal Processing and Integrated Networks (SPIN), 2014 International Conference on Date of Conference: 20-21 Feb. 2014 Page(s): 56 – 58.

[15] Lee, F. ; Massachusetts Institute of Technology, Cambridge, Mass "Reading machine: From text to speech",Published in: Audio and Electroacoustics, IEEE Transactions on (Volume:17 , Issue: 4) Page(s): 275 – 282.

[16] Liang Wang ; Dept. of Electron. Eng., Shanghai Jiao Tong Univ., Shanghai, China ; Jie Zhu ; Yao Lv "An improved method for predicting fundamental frequency contour in mandarin text-to-speech system with a small corpus" ,Published in: TENCON 2010 - 2010 IEEE Region 10 Conference Date of Conference: 21-24 Nov. 2010 Page(s): 751 – 754.

[17] Edgington, M. ; British Telecom Res. Labs., Ipswich, UK ; Lowry, A."Residual-based speech modification algorithms for text-to-speech synthesis" ,Published in: Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on (Volume:3) Date of Conference: 3-6 Oct 1996 Page(s): 1425 - 1428 vol.3.

[18] Xu Jun ; Inst. of Inf. Sci., Northern Jiaotong Univ., Beijing, China ; Yuan Baozong "Next generation of Chinese text-to-speech system" ,Published in: TENCON '93. Proceedings. Computer, Communication, Control and Power Engineering.1993 IEEE Region 10 Conference on (Volume:2) Date of Conference: 19-21 Oct. 1993 Page(s): 1078 - 1081 vol.2.

[19] Hamad, M. ; Electr. & Electron. Eng. Dept., Univ. of Khartoum, Khartoum, Sudan ; Hussain, M. "Arabic Text-To-Speech Synthesizer" ,Published in: Research and Development (SCOReD), 2011 IEEE Student Conference on Date of Conference: 19-20 Dec. 2011 Page(s): 409 – 414.

