

HMM Based TTS for Tamil

Dr.P. Prabakaran,

PDF(UGC), University Of Madras

Email: oliyavan@gmail.com

முன்னுரை

இயற்கை மொழி ஆய்வு இன்று உலகளவில் பல்வேறு மொழிகளுக்கு நடந்து கொண்டிருக்கிறது. அறிவியல் தொழில்நுட்ப வளர்ச்சியானது உலகளவில் மிக வேகமாக வளர்ந்து வருகிறது. மனித - இயந்திர உரையாடல் என்பது எவ்வாறு மனிதன் மனிதனுடன் உரையாடுகிறானோ அதேபோன்று கணினியுடன் உரையாடுமாறு அதனை தயார்படுத்துவதை உள்ளடக்கியது. இந்த வகையில் இயற்கை மொழி ஆய்வின் ஒரு பகுதியான பனுவல் - பேச்சு மாற்றியானது இன்று பல மொழிகளுக்கு உருவாக்கப்பட்டு வந்தாலும் தமிழுக்கென்று மிக துல்லியமான பனுவல் - பேச்சு மாற்றி இன்னும் உருவாக்கப்படவில்லை. அதற்குக் காரணம் தமிழ் மொழிக்கென்று ஒரு முழுமையான ஆய்வு மேற்கொள்ளப்படாததே ஆகும். கணினி மொழியியல் வல்லுநர்கள் கணித, புள்ளியல் அடிப்படையில் உள்ள கோட்பாடுகளை அறிவியல் தொழில்நுட்ப ரீதியில் ஆய்வுக்குட்படுத்தி மேல்நாட்டு வாய்ப்பாட்டினைத் தமது மொழிக்குப் புகுத்தினால் அன்றி நாம் ஒரு முழுமையான பனுவல் பேச்சு மாற்றியை உருவாக்க முடியாது. கணினித்தமிழ் வளர்ச்சியின் காரணமாக ஆங்கிலம் இன்று மிகப்பெரிய வளர்ச்சியை அடைந்துள்ளது. பல்வேறு சிறப்பம்சங்கள் கொண்ட மென்பொருட்கள் அம்மொழிக்கென உருவாக்கப்பட்டுள்ளது. மேலும் உருவாக்கப்படும் வருகிறது. அது போன்றதொரு வளர்ச்சியைத் தமிழுக்கும் சாத்தியப்படுத்துவது நம்முடைய கடமையாகிறது. பனுவல் - பேச்சு மாற்றிக்கென தமிழுக்கும் ஆங்காங்கே சில முயற்சிகள் மேற்கொள்ளப்பட்டு வருகிறது. ஆனால் அதன் துல்லியத்தன்மையை நம்மால் இன்னும் முழுமையாக அடைய முடியவில்லை என்பது மறுக்கமுடியாத உண்மையாகும். கணித மற்றும் புள்ளியல் அடிப்படையில் எழுத்து வடிவத்தை ஒலியன் வடிவமாக மாற்றி பிறகு ஒலியன் வடிவத்தை மாற்றொலியன்களாக மாற்றி மாற்றொலியன்களை ஒலி அலைகளாக மாற்றுவதன் மூலம் இதனை சாதிக்கலாம். HMM என்னும் கோட்பாட்டை எவ்வாறு நாம் தமிழ் மொழிக்குப் பயன்படுத்தலாம் என்பதே இக் கட்டுரையின் நோக்கமாகும்.

பனுவலிலிருந்து ஒலியன்களாக மாற்றுதல்

பனுவலிலிருந்து ஒலியன்களாக மாற்றுவதற்கு ஆங்கில மொழியின் அமைப்பானது மிகவும் எளிமையாகக் காணப்படுவதால் சிரமமின்றி அவற்றை ஒலியன்களாக மாற்ற முடிகிறது. ஆனால் தமிழ் மொழியின் அமைப்பானது பல

மாற்றொலிகளைக்(Allophones) கொண்டு அமைவதால் பனுவலிலிருந்து பேச்சொலிகளாக மாற்றுவதற்கு மூன்று நிலைகளைக் கையாள வேண்டியுள்ளது. முதலில் பனுவலிலிருந்து ஒலியன்களாக மாற்றி பிறகு ஒலியன்களை மாற்றொலிகளாக மாற்றி அந்த மாற்றொலிகளைப் பேச்சொலிகளாக மாற்ற வேண்டும். உதாரணமாக 'ச' என்னும் எழுத்து வார்த்தையின் முதலில் வரும்போது 'ச'(c) (சட்டை(Sattai) எனவும் இடையில் வரும்போது 'ச'(ch) பச்சை (pachchai) எனவும் மூக்கொலிக்கு அடுத்து வரும்போது 'ச'(ja) பஞ்சு எனவும் உச்சரிக்கப்படுவதால் இம்மூன்று நிலைகளை நாம் கையாள வேண்டியுள்ளது. இவ்வாறாக மாற்றப்பட்ட மாற்றொலிகளை நாம் ஒலி இயந்திரத்திற்கு உள்ளீடாக கொடுக்க வேண்டும். இது போன்ற தமிழ் பனுவலை மாற்றொலியன்களாக மாற்றுவதற்கு HMM மிகச்சிறந்த முறையில் ஆவணஞ் செய்கிறது.

வாய்ப்பாடு:

$$G = (g_1, g_2, g_3, \dots, g_N)$$

G என்பது பனுவலின் வரிசை ஆகும்.

$$Ph = (Ph_1, Ph_2, Ph_3, \dots, Ph_N)$$

Ph என்பது ஒலியனின் மறைந்த வரிசை ஆகும்.

$$W = \text{argmax}(P(G|Ph)P(Ph))$$

P(Ph) என்பது ஒலியனின் prior Probability sequence ஆகும்

மேலேயுள்ள பனுவல் குறியீடு G என்பது observation ஆகவும், Ph என்னும் ஒலியனின் குறியீடான மறைநிலை வரிசையாகவும் hidden state sequence ஆகவும் காணப்படுகிறது. பனுவல் – பேச்சொலியாக மாற்றுவதற்கு இந்த நிகழ்தகவு மிகவும் இன்றியமையாதது ஆகும்.

உதாரணமாக

'அம்மா' என்ற வார்த்தையை நாம் ஒலியன்களாக மாற்றும்போது தரவுகளில் பனுவலுக்கு நிகரான ஒலியன்களை நாம் கொடுக்கும் போது அந்த பனுவலுக்கு நிகரான ஒலியனை புள்ளியல் அடிப்படையில் மாற்றுகின்றன.

$$W = \text{argmax}(P(a|A)P(A))$$

இதுபோன்ற தரவுகளை நாம் உள்ளீடுகளாகக் கொடுக்கும் போது அனைத்து உள்ளீடுகளையும் புள்ளியியல் அடிப்படையில் ஒலியன்களாக மாற்றி அமைக்கின்றன.

Hidden Markov Model-ஐப் பயன்படுத்தி புள்ளியியலின் அடிப்படையில் எவ்வாறு ஒலியின் வேகத்தை(speech duration) கணக்கிட்டு பேச்சு ஒலியின் கால அளவையை(Time Duration), நிகழ்தகவு(probability) அடிப்படையில் ஆய்வுகூட்டுபடுத்தி பனுவல்-பேச்சு மாற்றியை உருவாக்க முடியும் என்பதை கீழ்க்கண்ட மூன்று நிலைகளில் காணலாம்

நிகழ்தகவு(Probability) என்பது மூன்று நிலைகளில் கணக்கிடப்படுகிறது

N = State

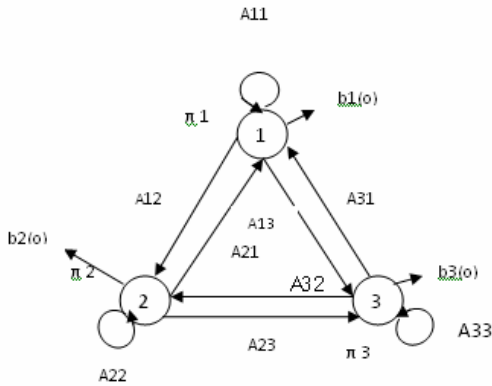
1 State Transition Probability $A = \{a_{ij}\}_{i,j=1}^N$

2 Output Probability $B = \{b_i(o)\}_{i=1}^N$

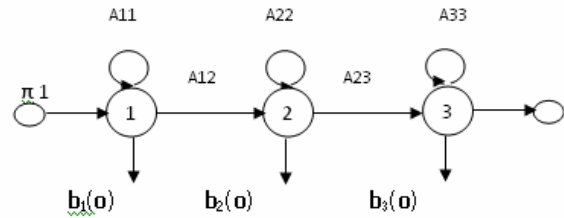
3 Initial state Probability $= \{\pi_i(0)\}_{i=1}^N$

இங்கு N என்பது இயல் எண் ஆகவும் A என்பது state transition ஆகவும் B என்பது முதல் நிலை நிகழ்கதவாகவும் காணப்படுகிறது. இங்கு உள்ளீடானது ஒரு நிலை(State) லிருந்து இன்னொரு நிலைக்கு(State) செல்லும்போது எவ்வாறு Markov's Model விதிபடி Hidden State ஆக மாறுகிறது என்பதைக் கீழ்க்காணும் படத்தின் மூலம் விளக்கலாம்.

A) 3 state Ergodic Model



B) A -3 state left to right model



ஒலியனிலிருந்து மாற்றொலிகளாக மாற்றுதல்

பனுவலிலிருந்து ஒலியனாக மாற்றி பிறகு ஒலியனிலிருந்து மாற்றொலிகளாக மாற்றும்போது ஒரு எழுத்தானது முந்தைய எழுத்தின் அரை ஒலி மற்றும் பிந்தைய எழுத்தின் அரை ஒலியைப் பெற்றிருக்கும். உதராணம் 'க' என்ற எழுத்து மூக்கொலிக்குப் பின்பு வந்தால் 'க' (g) எனவும் அதே எழுத்து வார்த்தைக்கு முதலில்

வந்தலால் 'க'(k) எனவும், அதே எழுத்து இரண்டு உயிரெழுத்துக்கு இடையில் வந்தால் 'க'(h) என உச்சரிக்க வேண்டும். இவ்வாறு தமிழ் மொழியிலுள்ள எல்லா ஒலியன்களுக்கும் சூழ்நிலையின் அடிப்படையில் கிட்டத்தட்ட 97 மாற்றொலிகளை உருவாக்க முடியும். இவ்வாறு உருவாக்கப்பட்ட மாற்றொலிகளை ஒலி இயந்திரத்திற்கு (sound Engine) உள்ளீடாகக் கொடுக்கும்போதுதான் நாம் இயற்கையான பேச்சொலியைப் பெறமுடியும்.

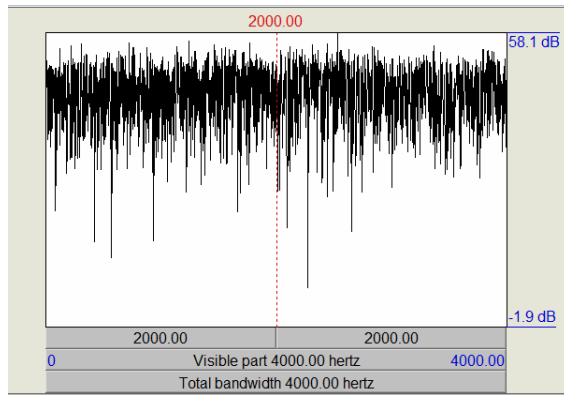
துல்லியத்தன்மையை அடைய மேற்கொள்ள வேண்டிய பணிகள்

நாம் பனுவலிலிருந்து பேச்சொலியாக மாற்றுவதற்கு மிக முக்கியமான ஆய்வாக இந்த **spectrum analysis** விளக்குகிறது. எடுத்து காட்டாக. அவனா என்ற வார்த்தையை இரு வகையாக நாம் உச்சரிக்கலாம். முதலில் அ) 'அவனா' என்று உச்சரிக்கும்போது யார் அவரா என பொருள்படும் அதுவே ஆ) 'அவனா' என உச்சரிக்கும் போது அவன் மோசமானவன் என பொருள்படும். எனவே இந்த இருவகையான உச்சரிப்பை எவ்வாறு இயந்திரம் உச்சரிக்க முடியும் என்பதை இவ்வகையான ஆய்வுகளை மேற்கொள்வதன் மூலம் செய்ய முடியும். கீழே கொடுக்கப்பட்டுள்ள படத்தின் மூலம் அவனா என்ற சொல் எவ்வாறு வேறுபடுகிறது என்பதைக் காணலாம்.

அ) Object type: Spectrum ,

Frequency domain: Lowest frequency: 0 Hz Highest frequency: 4000 Hz

Total bandwidth: 4000 Hz



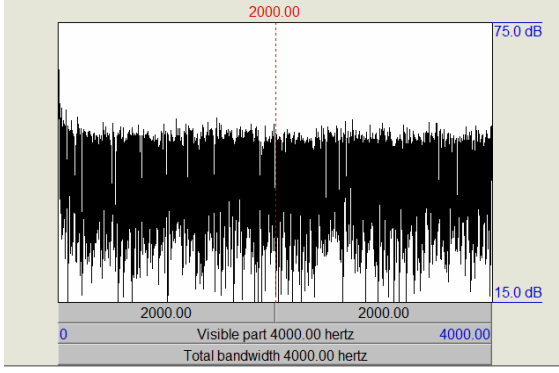
Frequency sampling: Number of frequency bands (bins): 2049 Frequency step (bin width): 1.953125 Hz First frequency band around (bin centre at): 0 Hz Total energy: 0.052854017 Pa² sec

ஆ) Object type: Spectrum , Frequency domain: Lowest frequency: 0 Hz

Highest frequency: 4000 Hz Total bandwidth: 4000 Hz

Frequency sampling: Number of frequency bands (bins): 16385

Frequency step (bin width): 0.244140625 Hz



First frequency band around
(bin centre at): 0 Hz)
Total energy: 0.254712705 Pa2 sec

முடிவுரை

பனுவல் பேச்சு மாற்றியானது இன்றைய காலக்கட்டத்தில் மிகவும் இன்றியமையாத ஒரு மென்பொருள் ஆகும். பல வகையான ஆய்வுகள் மேற்கொண்டாலும் குறிப்பாக மொழியியல் வல்லுநர்கள் இன்றி இதுபோன்ற ஒரு மென்பொருளை உருவாக்குவது என்பது சாத்தியமில்லாததாகும். இதுபோன்ற ஆய்வுகள் மூலம் ஒரு நல்ல பனுவல் பேச்சு மாற்றியை உருவாக்க முடியும். இதுபோன்ற மென்பொருள் துல்லியமாக உருவாக்கப்பட்டால் பார்வையிழந்த மாணவர்கள், படிப்பறிவு இல்லாதவர்கள் அடுத்தவர்களுடைய துணையின்றி தங்களுக்குத் தேவைப்படும் நூலைப் படித்து பயன்பெற முடியும்.

துணைநூற்பட்டியல்

- 1) [Daniel Jurafsky](#) (Author), [James H. Martin](#) (Author) , Speech and Language Processing, 2nd Edition Hardcover – May 16, 2008
- 2) X. D. Huang, Y. Ariki, and M. A. Jack, Hidden Markov models for speech recognition, Edinburgh University Press, 1990
- 3) K. Tokuda, T. Kobayashi, and S. Imai. Speech parameter generation from HMM using dynamic features. In Proc. ICASSP-95, pages 660–663, May 1995
- 4) T. Masuko, K. Tokuda, T. Kobayashi, and S. Imai. Speech synthesis using HMMs with dynamic features. In Proc. ICASSP-96 , pages 389– 392, May 1996.
- 5) <https://wiki.inf.ed.ac.uk/twiki/pub/CSTR/TrajectoryModelling/HTS-Introduction.pdf>