

## वाक् से पाठ तथा पाठ से वाक् परिवर्तक (संकर आधारित द्वि-भाषा मशीनी अनुवाद)

<sup>१</sup>कपिल विलासराव गावंडे, <sup>२</sup>प्रो. विजय कुमार कौल

<sup>१</sup>शोधार्थी, कपिल विलासराव गावंडे, सूचना एवं भाषा अभियांत्रिकी केंद्र, महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा, महाराष्ट्र, भारत.

<sup>२</sup>सेवानिवृत्त निदेशक, प्रो. विजय कुमार कौल, सूचना एवं भाषा अभियांत्रिकी केंद्र, महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा महाराष्ट्र, भारत.

Email – <sup>१</sup> kapil.gawande8@gmail.com, <sup>२</sup> kaulv2002@yahoo.com

**सारांश:** आधुनिकता के समय में मानव तथा भाषा के बीच के अंतरसंबंध को समझने में प्रकृतिक भाषा संसाधन की अहम भूमिका है इस अंतरसंबंध को समझने के लिए मशीन का उपयोग किया जाता है आज मानव भाषा को सीखने तथा बोलने के लिए मशीन का उपयोग कर रहा है इस शोध पत्र में एक भाषा वाक् को दूसरी भाषा वाक् में अनुवाद की विधि का उल्लेख किया है इस विधि को तीन चरणों में विभक्त किया है जिसमें ऑटोमैटिक वाक् संज्ञानक, मशीन अनुवाद और वाक् संश्लेषक शामिल है।

**मुख्य शब्द:** भाषा, मराठी, अंग्रेजी, मातृभाषा, वाक्, ऑटोमैटिक वाक् संज्ञानक, मशीन अनुवाद, वाक् संश्लेषण है।

### 1. परिचय:

स्रोत भाषा से लक्ष्य भाषा में रूपांतरण करना अनुवाद कहलाता है। वर्तमान समय में एक भाषा को दूसरी भाषा में समझने के लिए अनुवाद की भूमिका अत्यंत महत्वपूर्ण हो गई है। आज विश्व भर में अनुवाद की आवश्यकता को तीव्रता से महसूस किया जा रहा है। वैज्ञानिक टेक्नोलॉजी, कला-साहित्य, संस्कृति, दर्शन, राजनीति, समाज शास्त्र, वनस्पति विज्ञान आदि ज्ञान की तमाम शाखाओं में हो रहे एक भाषा के कार्य को दूसरे भाषा में अनुवाद की भूमिका महत्वपूर्ण है। मशीनी अनुवाद से तात्पर्य कंप्यूटर से है जिसके माध्यम से एक भाषा के पाठ को दूसरी भाषा में रूपांतरित किया जाता है, मशीनी अनुवाद कहलाता है। मशीनी अनुवाद के लिए स्रोत भाषा और लक्ष्य भाषा दोनों के ही वाक्यात्मक नियमों और शब्दकोश की आवश्यकता होती है। वाक्यात्मक नियमों को तार्किक अभिव्यक्ति तथा शब्दकोश को डेटाबेस के रूप में मशीन में संग्रहीत करते हुए मशीनी अनुवाद प्रणालियों को विकसित किया जाता है। वाक् से वाक् मशीन अनुवाद इसका मतलब यह है कि, “किसी एक भाषा के वाक् को बोलकर उसे मशीन द्वारा दूसरी भाषा के वाक् में अनुवाद करना वाक् से वाक् मशीन अनुवाद कहलाता है”।<sup>[7][6][2]</sup>

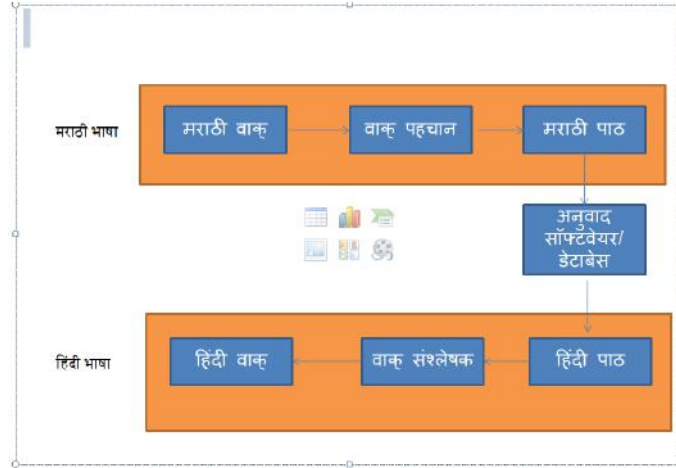
अब तक वाक् से वाक् मशीन अनुवाद पर कई सारे काम हुए हैं और अभी भी जारी है जैसे कि IBM, MASTOR जिसमें सीमित भाषाएँ हैं जिसका इस्तेमाल एंड्राइड मोबाइल के लिए किया जा रहा है, माइक्रोसॉफ्ट और SKYPE सॉफ्टवेयर आदि।

### 2. मशीन अनुवाद :

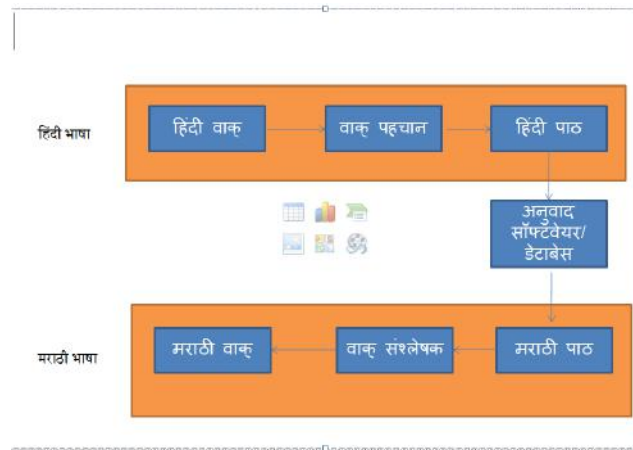
यह ऐसी सॉफ्टवेयर प्रणालियों के विकास से संबंधित है जिनके माध्यम से किसी एक प्राकृतिक भाषा के पाठ का अनुवाद दूसरी भाषा के पाठ में स्वचालित रूप से किया जा सके। मानव सहायता के साथ, कंप्यूटर या यंत्र का उपयोग कर एक प्राकृतिक मानव भाषा का दूसरी प्राकृतिक मानव भाषा में अनुवाद करना ही मशीन अनुवाद है। आज बाजार में कई सारी कंपनियाँ हैं जो मशीन अनुवाद के लिए साधन प्रदान करती हैं जैसे गूगल ट्रांसलेटर, माइक्रोसॉफ्ट ट्रांसलेटर आदि।<sup>[9]</sup>

#### 2.1 वाक् से वाक् की विकास हेतु प्रक्रिया

यह प्रणाली विकास हेतु ऑटोमैटिक वाक् संज्ञानक, मशीन अनुवाद, वाक् संश्लेषक तीन स्तरों पर आधारित है। इसके अंतर्गत मशीन अनुवाद से होने वाली गलतियों का सुधार होगा। इसके विकास की प्रक्रिया के लिए बनाई गई रेखांकित आकृति।



### 1.0 रेखांकित आकृति



### 1.1 रेखांकित आकृति

## 2.2. वाक् संज्ञानक पद्धतियों के प्रयोग:-

वाक् संज्ञानक पद्धति विकासशील है। जिसमें उच्चारण संबंधी ध्वनि की पहचान होती है यह पद्धति ऑनलाइन या ऑफलाइन तरीके से हो सकती है। ऑनलाइन में स्क्रिप्टिंग प्रोग्राम के इस्तेमाल से ध्वनि को पहचान कर पाठ में रूपांतरित करते हैं इसके इस्तेमाल से हम अंकन को पहचान कर प्रयोग में लाया जाता है।

वाक् संज्ञानक का निर्माण अन्य तकनीक के जरिए भी किया जाता है जिसमें ऑटोमैटिक वाक् संज्ञानक (ASR) को इस्तेमाल में लाया जाता है ASR के लिए बहुत बड़े कॉर्पस डेटा का इस्तेमाल किया जाता है जिससे की वाक् संज्ञानक आसानी से कार्य कर सके और त्रुटियों को रोका जा सके।[10]

## 2.3 अनुवाद पद्धतियों के प्रयोग

गूगल अनुवाद एपीआई का इस्तेमाल करके उसे पार्सिंग के जरिए पदच्छेदन कर पुनः वाक्य का पुनर्गठन करके गूगल अनुवाद से आने वाले पाठ को टोकन तथा सामान्यीकरण करके जांच कर उनमें सुधार किया जाएगा। जिसमें संकर (Hybrid) उपागम का प्रयोग कर व्याकरणिक एवं संदिग्ध शब्दों में रूल्स और अल्गोरिदम तथा शब्दकोश की सहायता के जरिये वाक्यों से त्रुटियों का सुधार होगा। गूगल अनुवाद से आने वाले छोटे से वाक्य को पदच्छेद करने हेतु पार्सिंग आवश्यक है ताकि बड़े पाठ को गठन कर अनुवाद कर सके।[2][4]

उदा. -

A एक शब्दों की श्रृंखला है जो गूगल अनुवाद होकर पार्सिंग की मदत से S स्ट्रिंग में परिवर्तित कर गठन होगा ताकि कितने भी बड़े पाठ का अनुवाद कर सके।

$$S = \sum_{A=0}^i A = A(i - 1)$$

तत्पश्चात आने वाले पाठ की जांच कर पाठ में रूल लगाकर वाक्य को सुधारा जाए।

उदाहरण -

S स्टिंग'जों की गूगल अनुवाद

S = (f)  $\sum_{A=0}^i A \{A = A(i - 1)\} - 1 \dots n$

S - 1.....n यह सुधारित त्रुटि के वर्ण या वाक्य है। जिसे रूल या वाक्य बदल कर लिख सकते हैं।

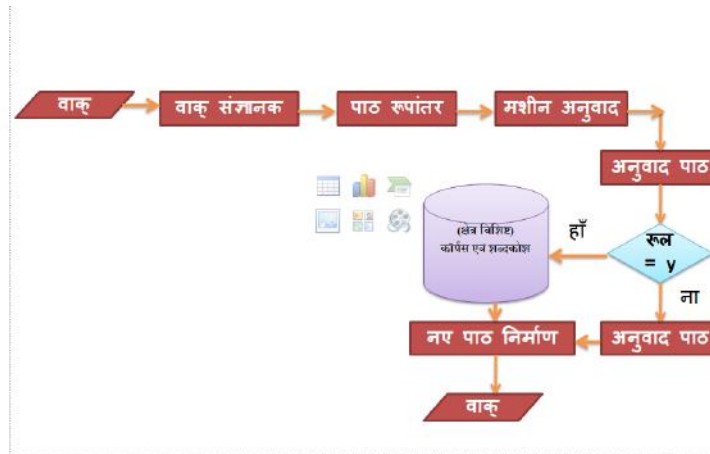
### 2.4 वाक् संश्लेषक पद्धतियों के प्रयोग

वाक् संश्लेषक पाठ की जांच कर उसे सही उच्चारण में परिवर्तित करता है। यह ध्वनि संकेत चिन्ह को पहचान कर उसे मशीन के जरिये बुलवाया जाता है। इसे TTS पद्धती या टेक्स्ट टु स्पीच भी कहते है। वाक् संश्लेषक हम अपने मनपसंदीदा व्यक्ति (पुरुष या स्त्री) की आवाज में मशीन या कंप्यूटर से बुलवा सकते हैं। यह एक बड़े कॉर्पस डेटाबेस से जुड़ा होता है जो की पाठ से कॉर्पस डेटाबेस की जांच कर पाठ का हर एक शब्द तथा वाक्य के अडियो फ़ाइल या .wav फ़ाइल से बुलवाया जाता है।

### 3. प्रयोगात्मक कार्य पद्धति:-

प्रोग्रामिंग कि सहायता से वाक् स्वरों को पहचान कर उसे पाठ में रूपांतरित करके उस पाठ को गूगल अनुवाद कि सहायता से पाठ को अनुवाद कर वह पाठ (क्षेत्र विशिष्ट) का है की जांच कर उसमें सुधार कर उसके बाद वाक् में प्राप्त किया गया है। हर एक स्तर की अच्छी तरह जांच कर कार्य में बदलाव किया गया है। कार्य पद्धती के हर एक पहलू को डेटा-फ़्लो आरेख द्वारा नीचे दिये गए आकृति के जरिये समझाया गया है।

#### 3.1 डाटा-फ़्लो आरेख



1.3 वाक् से वाक् कार्य पद्धति डाटा- फ़्लो आरेख

#### 3.2 अल्गोरिथम

स्टेप 1. Start

स्टेप2 . If वाक् != null

Check Speech API

Convert Speech API String S

Else

Return end if

स्टेप 3. Translation the word/ sentence

For create token in word

If token word = not correct (wrong)

Then check the rule and replace the word=S1

$$S = (f) \sum_{A=0}^i \{ A = A(i-1) - 1 \dots n \}$$

Else goto स्टेप 4

End if ,End for

स्टेप 4 . S1 = Speech Synthesizer

If

Speech Synthesizer = वाक्

Return (स्टेप 4) end if

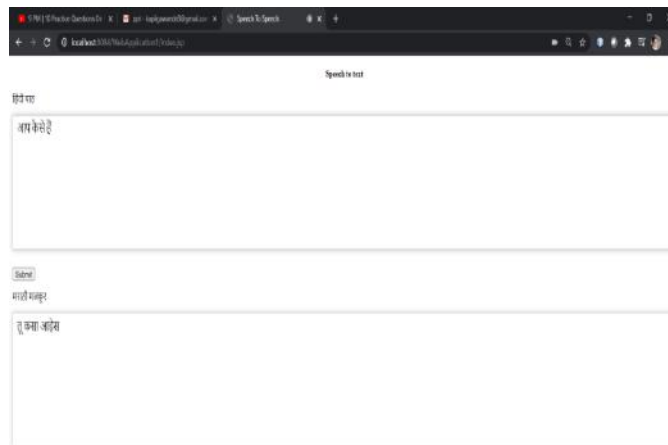
स्टेप 5. Stop/Exit

### 3.3 कार्यप्रणाली :

बोली गई हर वह स्रोत प्राकृतिक भाषा जिसका अर्थ निकलता हो वह वाक् कहलाती है। उपयोगकर्ता के बोले गए स्वरों की तरंग के जरिए मशीन में डिजिटल रूप में दर्ज करके मशीन में पहले से ही मौजूद बड़ा कॉर्पस या डेटाबेस से स्वरों की जांच कर वाक् पहचान करेगा। तत्पश्चात वाक् संज्ञानक के इस्तेमाल से वाक् को पाठ में परिवर्तित कर पहले से मौजूद मशीन अनुवाद जैसे गूगल अनुवाद, माइक्रोसॉफ्ट अनुवाद के इस्तेमाल से दिये गए पाठ को अनुवाद और उन्हे संघटित कर गंतव्य पाठ में परिवर्तन किया जाएगा। गंतव्य पाठ की त्रुटियों को रूल लगाकर या कार्पस एवं शब्दकोश से जांच कर उसमे सुधार कर नये पाठ का निर्माण होगा और वाक् संश्लेषण के इस्तेमाल से पाठ को वाक् में परिवर्तन कर पुनः वाक् को प्राकृतिक गंतव्य भाषा में मशीन के जरिये बोला जाएगा।

### 3.4 स्क्रीनशॉट :

हिंदी पाठ से मराठी पाठ:



### 1.4 मराठी पाठ से वाक् में परिवर्तन स्क्रीनशॉट

तू कसा आहेस।

### 3.5 डेटा विश्लेषण :

[14]सांख्यांकीय प्रतिदर्श विधि का उपयोग कर कुछ वाक्य को अनुवाद कर निम्न सूची निर्धारित की गई है-

भाषा स्रोत से गंतव्य	प्रतिशत	विकिपीडिया के अनुसार गूगल अनुवाद
हिंदी भाषा वाक्	६५	५०
हिंदी भाषा से मराठी पाठ	५५	५०
मराठी भाषा वाक्	६५	५०
मराठी से हिंदी पाठ	५०	५०
कुल प्रमाण	५८.७५%	५०%

### 4. निष्कर्ष :

वाक् से वाक् मशीन अनुवाद मराठी से हिंदी (क्षेत्र विशिष्ट) हिंदी एवं मराठी भाषाओं के अनुप्रयुक्त क्षेत्र के लिए लाभदायक होगा। आधुनिक वाक् से वाक् प्रक्रिया, यह मराठी और हिंदी द्वि-भाषा के लिए विकसित कर आगे बहुभाषी स्तर पर इस पर काम किया जाएगा। तथा बहुभाषी मशीन अनुवाद में उपयुक्त साबित होगा। वाक् से वाक् मशीन अनुवाद का प्रयोग प्राकृतिक भाषा संसाधन के विभिन्न क्षेत्रों में किया जा सकता है जैसे- विडियो अनुवाद, पाठशाला, गवर्नमेंट एंड नॉन गवर्नमेंट ऑफिस आदि।

### संदर्भ ग्रंथ – सूची :

1. Mahak Dureja, S. G. (November 2015). Speech-to-Speech Translation: A Review. International Journal of Computer Applications (0975 – 8887) , Volume 129 – No.13, Volume 129 – No.13.
2. Speechalator: two-way speech-to-speech translation on a consumer PDA. (2003). EUROSPEECH 2003 - GENEVA , 1-4.
3. Sudip Naskar, S. B. (2005). Use of Machine Translation in India: Current Status. MTS-2005 .
4. Vajna, S. (2002.). APPROACHES OF KNOWLEDGE-BASED DESIGN. INTERNATIONAL DESIGN CONFERENCE - DESIGN 2002, Dubrovnik , 1-8.
5. डॉ. रविद्रनाथ श्रीवास्तव, ड. क. अनुवाद सिद्धांत और समस्याएँ.
6. तिवारी, ड. भ. (१९८४). आधुनिक भाषा विज्ञान. इलाहबाद: किताब महल प्राइवेट लिमिटेड.
7. प्रसाद, ध. (२०१६). भाषाविज्ञान का सैद्धांतिक अनुप्रयुक्त एवं तकनीकी पक्ष.
8. स्रोत:<<https://pdfs.semanticscholar.org/0fa1/911622a6c0a3dd43fefbdf2695ebdb7e10fa.pdf> >
9. स्रोत:<<http://www.aclweb.org/anthology/W06-3708> >
10. Farzad Ehsani, Jim Kimzey, Demitrios Master, Karen Sudre, David Domingo Hunil Park Engineering Department Sehda, Inc.
11. स्रोत:<<http://prakashblog-google.blogspot.com/2010/09/1-introduction-of-machine-translation.html>>
12. स्रोत:<[https://lgandlt.blogspot.com/2016/11/blog-post\\_13.html?m=0](https://lgandlt.blogspot.com/2016/11/blog-post_13.html?m=0)>
13. स्रोत:<[https://www.cdac.in/index.aspx?id=pdf\\_annualreport13-14\\_hindi](https://www.cdac.in/index.aspx?id=pdf_annualreport13-14_hindi)>
14. स्रोत <[https://en.wikipedia.org/wiki/Google\\_Translate](https://en.wikipedia.org/wiki/Google_Translate)>