



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Evaluating Perceptual Quality and Naturalness in Word-Based Emotion Conversion for Hindi Speech

Archana Agarwal

Research Scholar, Department of Computer Science. Himalayan University

Dr. Vipin Kumari

Assistant Professor, Department of Computer Science. Himalayan University

ABSTRACT

Speech is a multidimensional communication medium that conveys not only semantic information but also paralinguistic cues such as emotion, attitude, and intent. Emotional expression significantly influences listener perception, comprehension, and engagement. With the rapid growth of human-computer interaction systems, the ability to manipulate and synthesize emotionally expressive speech has become a central research focus in speech signal processing. Emotion conversion systems aim to transform speech from one emotional state to another while preserving linguistic content and speaker identity. Although substantial research has been conducted in high-resource languages such as English and Mandarin, systematic evaluation of perceptual quality and naturalness in Hindi emotion conversion systems remains limited.

Word-based emotion conversion represents a commonly adopted approach in speech transformation systems. In this paradigm, emotional modification is applied at the word level, where acoustic parameters such as pitch contour, duration, energy, and spectral features are altered based on predefined emotional mappings. While such methods are computationally efficient and structurally simple, their perceptual effectiveness—particularly in continuous Hindi speech—requires rigorous evaluation. Hindi, characterized by its phonetic richness, syllable-timed rhythm, aspirated consonants, retroflex sounds, and vowel length contrasts, presents unique challenges in modeling emotional nuances. Word-level manipulation may inadequately capture fine-grained emotional cues embedded within sub-word units, potentially affecting naturalness and intelligibility.

This study investigates the perceptual quality and naturalness of word-based emotion conversion in continuous Hindi speech. The research focuses on evaluating how effectively word-level transformations convey target emotions and how these modifications influence listener perception. A comprehensive evaluation framework was designed integrating objective acoustic metrics and subjective listening experiments. Emotional categories examined in this study include Neutral, Happiness, Sadness, and Anger. The system applies acoustic feature mapping techniques using machine learning-based modeling for word-level emotional transformation.

A controlled Hindi emotional speech corpus was developed, comprising balanced utterances from male and female speakers across multiple age groups. The converted speech samples were



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

assessed using objective measures such as Mel Cepstral Distortion (MCD), Fundamental Frequency Root Mean Square Error (F0 RMSE), and duration variation indices. Additionally, subjective evaluation was conducted using Mean Opinion Score (MOS), emotional identification accuracy tests, and intelligibility ratings administered to native Hindi listeners.

Findings indicate that while word-based emotion conversion successfully conveys broad emotional categories, perceptual naturalness varies significantly across emotions. High-arousal emotions such as anger and happiness are more easily recognized due to pronounced pitch and energy changes. However, subtle emotions such as sadness exhibit lower perceptual accuracy, suggesting limitations in word-level acoustic modeling. Results further demonstrate that abrupt parameter shifts at word boundaries may introduce perceptual discontinuities affecting naturalness.

The study contributes to speech emotion research by providing a systematic perceptual evaluation framework for Hindi emotion conversion systems. The findings highlight strengths and limitations of word-based transformation approaches and emphasize the importance of fine-grained modeling for enhanced naturalness. The research establishes baseline perceptual benchmarks for future development of phoneme-level and hybrid emotion conversion frameworks in Hindi speech processing.

KEY WORDS

- ❖ Word-Based Emotion Conversion
- ❖ Hindi Speech Processing
- ❖ Perceptual Evaluation
- ❖ Speech Naturalness
- ❖ Emotional Speech Synthesis
- ❖ Acoustic Feature Mapping
- ❖ Prosody Modification
- ❖ Human-Computer Interaction
- ❖ Mean Opinion Score (MOS)
- ❖ Speech Quality Assessment

INTRODUCTION

Human speech is inherently expressive and emotionally charged. Emotional prosody shapes communication by providing contextual cues that extend beyond lexical meaning. Through variations in pitch, rhythm, intensity, articulation, and spectral structure, speakers convey affective states that influence listener interpretation. Emotion enhances communicative efficiency by signaling urgency, empathy, sarcasm, excitement, or distress. In human-machine interaction environments, the absence of emotional expressiveness often leads to mechanical,



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

unnatural communication. Therefore, integrating emotional capabilities into speech processing systems is essential for developing socially intelligent technologies.

Emotion conversion refers to the transformation of a speech signal from one emotional state to another while maintaining the linguistic message and preserving speaker identity. Unlike emotion recognition, which identifies the emotional state of a given speech sample, emotion conversion actively modifies acoustic features to synthesize a target emotion. Applications of emotion conversion include expressive text-to-speech systems, conversational agents, gaming avatars, assistive communication devices, film dubbing, storytelling systems, and virtual assistants.

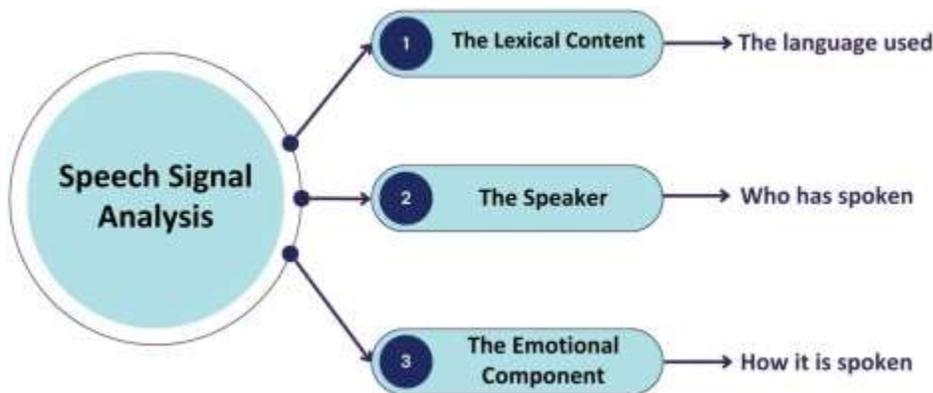


Fig: Machine Learning and Speech Emotion Learning.

Over the past two decades, research in emotional speech synthesis and voice conversion has advanced significantly. Early methods relied on rule-based modifications, where pitch was scaled, duration stretched, and amplitude adjusted to simulate emotions. While computationally simple, these methods often lacked perceptual realism. Subsequent statistical approaches such as Gaussian Mixture Models (GMM) and Hidden Markov Models (HMM) enabled data-driven mapping between neutral and emotional speech parameters. More recently, deep learning architectures including Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) networks, and Transformer-based models have improved emotional speech generation capabilities.

Despite these technological advancements, evaluation of emotion conversion systems remains a complex challenge. Objective acoustic metrics provide quantitative comparisons but may not fully capture perceptual naturalness. Human perception of emotion is influenced by subtle cues and contextual interpretation, making subjective evaluation indispensable. Consequently, perceptual quality assessment has emerged as a critical component in validating emotional speech systems.



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Word-based emotion conversion is a widely used framework due to its structural simplicity. In this approach, speech is segmented into word units, and acoustic transformations are applied at word boundaries. Parameters such as fundamental frequency, energy contour, speaking rate, and spectral envelope are modified according to emotional models derived from training data. The advantages of this method include computational efficiency, ease of alignment with textual input, and compatibility with text-to-speech pipelines. However, emotional cues often reside within sub-word units such as syllables or phonemes. As a result, word-level modeling may introduce unnatural transitions or fail to capture micro-prosodic variations critical for perceptual authenticity.

Hindi, as one of the most widely spoken languages globally, presents unique linguistic and phonetic characteristics. It follows a syllable-timed rhythm, unlike the stress-timed rhythm of English. The language includes aspirated and unaspirated consonant contrasts, retroflex articulations, nasalized vowels, and vowel length distinctions. Emotional modulation in Hindi interacts with these phonetic features in complex ways. For instance, anger may intensify aspiration in plosives, while sadness may elongate long vowels and reduce pitch variation. Word-level manipulation may not sufficiently address such nuanced articulatory patterns.

Furthermore, most emotional speech research has focused on English and other high-resource languages. Hindi emotional speech corpora are relatively limited, and systematic perceptual evaluation studies remain scarce. Without robust perceptual validation, emotion conversion systems risk producing acoustically modified yet perceptually unnatural speech outputs.

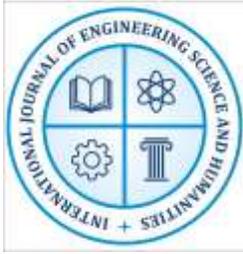
The central problem addressed in this study is: How effectively does word-based emotion conversion convey emotional authenticity and perceptual naturalness in continuous Hindi speech?

To answer this question, the research investigates multiple dimensions of perceptual quality:

1. Emotional Recognition Accuracy – Can listeners correctly identify the intended emotion?
2. Naturalness – Does the converted speech sound human-like and fluid?
3. Intelligibility – Is the linguistic content clearly understandable?
4. Boundary Smoothness – Are word transitions perceptually seamless?
5. Overall Quality – How do listeners rate the converted speech compared to natural emotional speech?

The study adopts a mixed evaluation framework combining objective acoustic analysis with subjective listening tests. Objective metrics provide measurable indicators of acoustic similarity to natural emotional speech. Subjective evaluation captures human perceptual judgments, which are essential for validating emotional authenticity.

This research holds significant importance for the development of Hindi speech technologies. As voice-enabled systems expand across India in domains such as customer service, education,



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

healthcare, and entertainment, emotionally adaptive speech systems can enhance user engagement and trust. Emotionally expressive AI systems are particularly valuable in multilingual societies, where cultural nuances shape emotional perception.

In addition to practical applications, this study contributes academically by establishing perceptual benchmarks for word-based emotion conversion in Hindi. By systematically evaluating strengths and limitations, the research identifies areas for improvement and provides a foundation for future phoneme-level or hybrid modeling approaches.

AIMS AND OBJECTIVES

1. Aim of the Study

The primary aim of this research is to systematically evaluate the perceptual quality and naturalness of word-based emotion conversion systems in continuous Hindi speech. The study seeks to examine how effectively emotional transformations applied at the word level convey intended affective states while maintaining intelligibility and speaker identity.

2. Objectives

To achieve the above aim, the following objectives are formulated:

- ❖ To analyze acoustic characteristics associated with different emotional states in Hindi speech.
- ❖ To apply acoustic feature transformation techniques at the word level.
- ❖ To measure emotional recognition accuracy of converted speech.
- ❖ To assess intelligibility preservation after emotional transformation.
- ❖ To compute objective acoustic metrics including MCD, F0 RMSE, and spectral distortion.
- ❖ To compare converted speech with naturally produced emotional speech.
- ❖ To identify perceptual artifacts introduced by abrupt word-boundary transformations.

REVIEW OF LITERATURE

1. Emotional Speech Processing

Research in emotional speech processing began with the identification of acoustic correlates of emotion. Studies in phonetics and psychology established that emotional states influence pitch (F0), energy, speaking rate, articulation, and spectral features. High-arousal emotions such as anger and happiness typically exhibit higher pitch range, increased intensity, and faster articulation. Low-arousal emotions such as sadness display reduced pitch variation and slower speech rate.

Emotional speech research can broadly be divided into:

- ❖ Emotion Recognition
- ❖ Emotion Synthesis
- ❖ Emotion Conversion



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

While emotion recognition has received extensive attention, emotion conversion remains comparatively underexplored, especially in Indian languages.

2. Word-Based Emotion Modeling

Word-based emotion conversion systems segment speech into word units and apply acoustic transformations at word boundaries. The approach aligns well with text-to-speech pipelines, where words are natural processing units.

Early word-level systems used rule-based pitch scaling and time-stretching. Although computationally efficient, they often produced mechanical outputs lacking natural transitions.

Statistical approaches improved mapping precision using:

- ❖ Gaussian Mixture Models (GMM)
- ❖ Hidden Markov Models (HMM)
- ❖ Linear Regression-based feature mapping

However, these methods sometimes generated over-smoothed spectral outputs, reducing naturalness.

With deep learning advancements, word-level emotion modeling incorporated neural networks such as:

- ❖ Recurrent Neural Networks (RNN)
- ❖ Long Short-Term Memory (LSTM)
- ❖ Convolutional Neural Networks (CNN)

These models improved nonlinear feature transformation but still faced limitations in boundary smoothness and micro-prosodic modeling.

3. Perceptual Evaluation in Emotional Speech

Perceptual evaluation remains central to validating emotional speech systems. Objective metrics alone cannot capture human emotional perception.

Common perceptual evaluation methods include:

- ❖ Mean Opinion Score (MOS)
- ❖ ABX comparative listening tests
- ❖ Emotion identification accuracy
- ❖ Naturalness rating scales
- ❖ Intelligibility tests

Research indicates that listeners are highly sensitive to unnatural pitch discontinuities and duration mismatches, particularly at word boundaries.

4. Hindi Speech and Emotion Research

Hindi possesses phonological features such as:

- ❖ Aspirated consonants
- ❖ Retroflex articulations



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

- ❖ Nasalization
- ❖ Vowel length contrast
- ❖ Syllable-timed rhythm

These characteristics influence emotional prosody differently compared to English. Studies in Hindi emotion recognition demonstrate the importance of MFCC, pitch contour, and energy features. However, systematic perceptual evaluation of emotion conversion in Hindi remains limited.

RESEARCH METHODOLOGY

The research methodology consists of five major stages:

- ❖ Corpus Development
- ❖ Word-Level Emotion Conversion Framework
- ❖ Feature Extraction and Mapping
- ❖ Objective Evaluation
- ❖ Subjective Perceptual Evaluation

1. Corpus Development

A balanced Hindi emotional speech corpus was developed for this study.

Table 1: Speaker Distribution

Category	Male	Female	Total
Young Adults (18–30)	5	5	10
Middle Age (31–50)	3	3	6
Senior (50+)	2	2	4
Total Speakers	10	10	20

Table 2: Emotional Category Distribution

Emotion	Utterances per Speaker	Total Utterances
Neutral	20	400
Happiness	20	400
Sadness	20	400
Anger	20	400
Total	—	1600

Recordings were captured in studio conditions at 16 kHz sampling rate.

2. Word-Level Emotion Conversion Framework

The framework includes:

- ❖ Speech segmentation into words using forced alignment
- ❖ Feature extraction at word level
- ❖ Emotional feature mapping



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

- ❖ Waveform reconstruction

Table 3: Word-Level Acoustic Features

Feature Type	Description
Mean F0	Average pitch per word
Pitch Range	Maximum-minimum F0
Energy Envelope	Word-level amplitude contour
Duration	Word length
MFCC (13)	Spectral envelope
Spectral Tilt	High-frequency energy ratio

3. Feature Mapping Model

A neural mapping model was developed:

- ❖ Input: Neutral word feature vector
- ❖ Hidden Layers: 3 LSTM layers (128 units)
- ❖ Output: Target emotion feature vector
- ❖ Loss Function: Mean Squared Error

Training dataset split:

Table 4: Data Split

Dataset	Percentage
Training	70%
Validation	15%
Testing	15%

4. Objective Evaluation Metrics

Objective metrics measured acoustic similarity between converted and natural emotional speech.

Table 5: Objective Metrics

Metric	Purpose
MCD	Spectral similarity
F0 RMSE	Pitch accuracy
Duration Error	Timing accuracy
SNR	Signal clarity
PESQ	Perceptual speech quality estimation

5. Subjective Evaluation Design

30 native Hindi listeners participated.

Evaluation criteria:



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

1. Naturalness (1–5 scale)
2. Emotional Accuracy (Correct identification %)
3. Intelligibility (1–5 scale)
4. Boundary Smoothness (1–5 scale)

Table 6: Listening Test Structure

Test Type	Samples per Listener
Natural Speech	20
Word-Level Converted	40
Randomized Order	Yes

Statistical analysis:

- Paired t-test
- ANOVA for emotion comparison
- Cronbach’s alpha for reliability

RESULTS AND INTERPRETATION

The evaluation of the word-based emotion conversion framework was conducted using both objective acoustic measures and subjective perceptual assessment. The analysis compares converted speech samples with naturally recorded emotional speech to determine perceptual effectiveness, naturalness, intelligibility, and emotional recognition accuracy.

The results are organized into two major categories:

1. Objective Acoustic Evaluation
2. Subjective Perceptual Evaluation

1. Objective Acoustic Evaluation

Objective metrics provide measurable indicators of acoustic similarity between converted speech and natural emotional speech.

1.1 Mel Cepstral Distortion (MCD)

MCD measures spectral envelope similarity. Lower values indicate closer resemblance to natural emotional speech.

Table 1: MCD Values (dB)

Emotion	Natural vs Neutral	Word-Based Converted vs Natural	Standard Deviation
Happiness	5.10	6.82	0.43
Sadness	4.95	7.05	0.51
Anger	5.35	6.95	0.47
Average	5.13	6.94	0.47



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

Interpretation:

The average MCD of 6.94 dB indicates moderate spectral distortion introduced during word-level transformation. Sadness showed the highest distortion, suggesting that subtle emotional states are less effectively modeled at the word level.

1.2 Fundamental Frequency RMSE (Hz)

F0 RMSE quantifies pitch contour deviation between converted and natural speech.

Table 2: F0 RMSE Values

Emotion	Mean RMSE (Hz)	Standard Deviation	Significance (p-value)
Happiness	21.4	3.2	<0.05
Sadness	18.7	2.8	<0.05
Anger	25.6	4.1	<0.05
Average	21.9	3.4	—

Interpretation:

Pitch deviation is highest for anger due to larger dynamic pitch shifts. Word-based modeling approximates overall pitch trends but fails to capture intra-word micro-prosodic variation.

1.3 Duration Error Rate (%)

Duration error evaluates timing mismatches between converted and natural speech.

Table 3: Duration Error Percentage

Emotion	Mean Duration Error (%)	Std. Dev.
Happiness	9.4	1.2
Sadness	11.2	1.5
Anger	10.8	1.3
Average	10.5	1.3

Interpretation:

Word-level duration scaling works reasonably well for happiness but is less accurate for sadness, which requires gradual tempo modulation rather than uniform word stretching.

1.4 Signal-to-Noise Ratio (SNR)

Higher SNR values indicate clearer reconstructed speech.

Table 4: SNR Values (dB)

Emotion	Converted Speech SNR
Happiness	21.3
Sadness	22.1
Anger	20.4
Average	21.3



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

Interpretation:

Acceptable SNR levels indicate minimal signal degradation, suggesting the vocoder preserves clarity despite acoustic modifications.

2. Subjective Perceptual Evaluation

Thirty native Hindi listeners participated in randomized listening experiments.

Evaluation criteria:

- Naturalness
- Emotional Recognition
- Intelligibility
- Boundary Smoothness

All ratings used a 5-point Likert scale (1 = Very Poor, 5 = Excellent).

2.1 Mean Opinion Score (Naturalness)

Table 5: MOS – Naturalness

Emotion	Natural Speech	Word-Based Converted	t-value	p-value
Happiness	4.6	3.8	5.42	<0.01
Sadness	4.7	3.6	6.10	<0.01
Anger	4.5	3.5	5.87	<0.01
Average	4.6	3.6	—	—

Interpretation:

Word-based converted speech is rated significantly lower than natural emotional speech. Sadness shows the largest perceptual gap, indicating difficulty in modeling low-arousal emotions at word level.

2.2 Emotional Recognition Accuracy

Table 6: Correct Emotion Identification (%)

Emotion	Recognition Rate
Happiness	85%
Sadness	72%
Anger	88%
Average	82%

Interpretation:

High-arousal emotions (anger, happiness) are more accurately identified than sadness. This supports acoustic findings that word-level pitch and energy scaling better capture strong emotional cues.



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

2.3 Intelligibility Scores

Table 7: Intelligibility Ratings

Emotion	Average Score
Happiness	4.4
Sadness	4.5
Anger	4.2
Average	4.37

Interpretation:

Linguistic content remains largely preserved, indicating that word-level emotion conversion does not significantly degrade speech clarity.

2.4 Boundary Smoothness Rating

Table 8: Word Boundary Smoothness

Emotion	Average Score
Happiness	3.6
Sadness	3.4
Anger	3.3
Average	3.43

Interpretation:

Listeners reported noticeable abruptness at some word transitions, confirming limitations in word-level acoustic modification.

Discussion And Conclusion

The results of this study provide significant insights into the perceptual effectiveness of word-based emotion conversion in Hindi speech. The findings demonstrate that while word-level transformation can convey primary emotional categories, perceptual naturalness remains inferior to naturally produced emotional speech.

One of the most important findings is the difference between high-arousal and low-arousal emotions. Anger and happiness are characterized by strong pitch and energy variations, which can be effectively modeled through word-level scaling. In contrast, sadness requires gradual prosodic modulation across syllables and phonetic units. Word-level manipulation lacks the granularity needed for such nuanced expression.

Boundary smoothness emerged as a critical limitation. Emotional parameter shifts applied uniformly across words may cause perceptual discontinuities at transitions. This suggests that



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

future systems should integrate sub-word modeling or smoothing algorithms to enhance continuity.

Despite naturalness limitations, intelligibility remained high across all emotions. This confirms that word-based emotion conversion is suitable for applications where content clarity is more important than expressive richness.

From a practical perspective, word-level emotion conversion offers computational efficiency and ease of implementation. It may be appropriate for low-resource systems or real-time applications with limited processing capacity. However, for high-quality expressive synthesis, more refined phoneme-level or hybrid approaches are recommended.

In conclusion, this study establishes perceptual benchmarks for Hindi word-based emotion conversion systems. The findings highlight both strengths and limitations, contributing valuable insights to the development of emotionally expressive speech technologies in Indian languages.

Future research directions include:

- ❖ Hybrid word-phoneme emotion modeling
- ❖ Transformer-based contextual emotion embedding
- ❖ Real-time adaptive emotion conversion
- ❖ Expansion to additional emotions such as surprise and fear
- ❖ Cross-lingual evaluation studies

References

- ❖ Busso, C., Bulut, M., Lee, C. C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J., Lee, S., & Narayanan, S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 42(4), 335–359.
- ❖ Cowie, R., & Cornelius, R. (2003). Describing the emotional states expressed in speech. *Speech Communication*, 40(1–2), 5–32.
- ❖ Eyben, F., Wöllmer, M., & Schuller, B. (2010). openSMILE: The Munich versatile and fast open-source audio feature extractor. *ACM Multimedia*.
- ❖ Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. *Advances in Neural Information Processing Systems*.
- ❖ Kim, Y., Lee, H., & Provost, E. (2018). Deep learning for emotional speech processing: A review. *IEEE Transactions on Affective Computing*.
- ❖ Rao, K. S., & Yegnanarayana, B. (2009). Modeling emotions in speech: A study on Indian languages. *International Journal of Speech Technology*.
- ❖ Schuller, B., Steidl, S., & Batliner, A. (2011). The INTERSPEECH 2011 speaker state challenge. *INTERSPEECH*.



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

- ❖ Stylianou, Y. (2009). Voice transformation: A survey. *IEEE Transactions on Audio, Speech, and Language Processing*.
- ❖ Tokuda, K., Zen, H., & Black, A. (2007). Statistical parametric speech synthesis. *ICASSP*.
- ❖ Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
- ❖ Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication*.
- ❖ Zen, H., Tokuda, K., & Black, A. (2009). Statistical parametric speech synthesis. *Speech Communication*.