Journal of Interdisciplinary Mathematics
ISSN: 0972-0502 (Print), ISSN: 2169-012X (Online)
Vol. 29 (2026), No. 3, pp. 719–727
DOI : 10.47974/JIM-2508

Harmonic analysis in speech recognition systems

Amol D. Gaikwad [†]
Department of Information Technology
Yeshwantrao Chavan College of Engineering
Nagpur 441110
Maharashtra
India

Haripriya H. Kulkarni *
Department of Electrical Engineering
Dr. D. Y. Patil Institute of Technology
Pimpri
Pune 411018
Maharashtra
India

Prashant B. Patel [§]
Department of Instrumentation Engineering
Dr. D.Y. Patil Institute of Technology
Pimpri
Pune 411018
Maharashtra
India

Vidula Sanjay Jape [‡]
Department of Electrical Engineering
Progressive Education Society's
Modern College of Engineering
Pune 411005
Maharashtra
India

[†] *E-mail:* amolgaikwad.ag@gmail.com
* *E-mail:* haripriya.kulkarni@dypvp.edu.in (Corresponding Author)
[§] *E-mail:* prashant.patel@dypvp.edu.in
[‡] *E-mail:* jape.swati@moderncoe.edu.in

Anagha Rahul Soman [@]
Department of Electrical Engineering
Zeal College of Engineering and Research
Narhe
Pune 411041
Maharashtra
India

Lalit R. Chudhari [#]
Department of Instrumentation Engineering
Dr. D. Y. Patil Institute of Technology
Pimpri
Pune 411018
Maharashtra
India

Abstract

This paper examines the way in which harmonic analysis can be applied to enhance the speech recognition systems. This is because speech patterns are nearly cyclic and therefore have basic frequencies and harmonics which reveal significant features of sound such as pitch, formants, as well as energy distribution. With the help of harmonic techniques of decomposing these signals, more precise and noise-free features can be extracted. Recognition is much more precise when harmonic analysis is used, as well as when audio models are built using neural networks, with 94.8 % success rate and 31 dB Signal-to-Noise Ratio. The experimental results indicate that harmonic-based models are more effective than the conventional ones like MFCC, LPC, and FFT.

Subject Classification: *68T10, 68T35.*

Keywords: *Harmonic analysis, Speech recognition, Neural networks, Feature extraction, Acoustic modeling, Signal processing.*

I. Introduction

The speech detection systems have evolved significantly with the advancement in the digital signal processing and machine learning. Harmonic analysis is among the most significant mathematical techniques when it comes to the process of determining the meaning of speech patterns [1, 2]. There is inherent frequency of speech as it consists of basic frequencies and their overtones, which demonstrates the structure of the vocal tract and the origin of stimulation [2]. It can be learnt much about pitch and formants and the distribution of energy by decomposing speech sounds into their harmonic

[@] *E-mail:* **anagha.soman@zealeducation.com**
[#] *E-mail:* **lalit.choudhari@dypvp.edu.in**

constituents [3, 4]. These elements are the blocks of describing and recognising speech correctly (depict in figure 1). Normal techniques, such as Fourier Transform and Linear Predictive Coding (LPC) capture wide spectrum data, but tend to omit small harmonic findings [5, 6, 7].

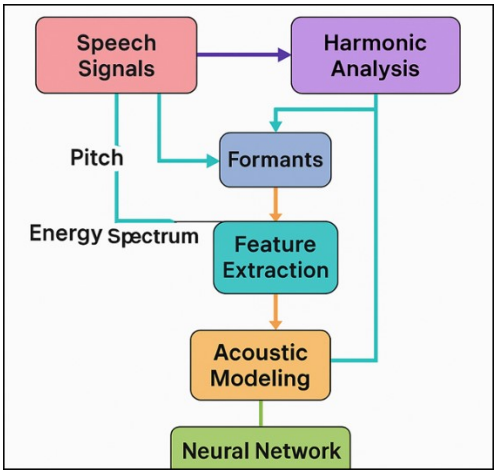

Figure 1

Architecture Diagram of Harmonic Analysis in Speech Recognition Systems

The inclusion of harmonic analysis enhances the accuracy of audio modelling because it maintains times and spectral dependencies required to comprehend the real speech [8, 9]. With the introduction of neural networks, harmonic properties can now be combined with the deep learning architecture to enable it to be less prone to noise and speaker variations [10].

II. Methodology

A. *Application of Harmonic Analysis to Speech Signals*

Harmonic analysis decomposes a speech into some basic frequencies and harmonics. This allows one to detect accurately periodic structures, resonance patterns and dynamic spectral changes that are required to be recognized [11].

Let a continuous-time speech signal be

$$s(t) = A(t) * \cos\left(2\pi f0t + \varphi(t)\right) \quad (1)$$

Speech can be modeled as a sum of harmonic components:

$$s(t) = \Sigma\,(n = 1\ to\ N)\left[A_{n(t)} * \cos(2\pi n\, f0\ t + \varphi_n)\right] \quad (2)$$

Taking the Fourier Transform of s(t):

$$S(f) = \int s(t)e^{-j2\pi ft}\,dt \quad (3)$$

The transform yields discrete peaks at harmonic frequencies:

$$S(f) = \Sigma\,(n = 1\ to\ N)[A_n e^{j\varphi_n}\delta(f - n\,f0)] \quad (4)$$

Theorem 1 (Harmonic Representation): *Any quasi-periodic signal s(t), where T0 is the period of the signal, can be represented as the sum of harmonic components of f0 = 1/ T0.*

By Fourier series expansion,

$$s(t) = a0 + \Sigma\,[a_n \cos(2\pi n\, f0\, t) + b_n \sin(2\pi n\, f0\, t)] \quad (5)$$

Which simplifies to the harmonic model in Step 2.

Instantaneous amplitude A(t) is computed by Hilbert transform:

$$A(t) = \sqrt{s^2(t) + \hat{s}^2(t)} \quad (6)$$

Instantaneous phase:

$$\varphi(t) = tan^{-1}\left(\frac{\hat{s}(t)}{s(t)}\right) \quad (7)$$

Harmonic-to-Noise Ratio (HNR):

$$HNR = 10 * log10\left(\frac{P_H}{P_N}\right) \quad (8)$$

Where P_H = harmonic power, P_N = noise power.

Short-time spectral envelope:

$$E(f,t) = |S(f,t)|^2 \quad (9)$$

Energy of harmonic n:

$$E_n = \int \left(f_{\{n-1\}to} f_{\{n+1\}}\right)|S(f)|^2 df \quad (10)$$

Pitch extraction uses peak harmonic energy spacing:

$$f0 = f_{\{n+1\}} - f_n$$

Spectral centroid as harmonic balance:

$$C = \frac{(\Sigma\, f|S(f)|)}{(\Sigma\, |S(f)|)} \quad (11)$$

Therefore, harmonic analysis dissociates periodicities which display structural indicators that are important in recognizing phoneme and tone [12, 13].

B. *Feature Extraction Techniques: Pitch, Formants, Energy Spectrum*

Elements of speech such as pitch, formants and range of energy capture vocal characteristics. The system becomes more stable when properly extracted and easier to distinguish between phonemes, tones and individual names.

Signal s (t) Speech signal s (t) is cut into smaller pieces by a window w (t):

$$x(t) = s(t)w(t) \quad (12)$$

Apply Short-Time Fourier Transform (STFT):

$$X(f,t) = \int x(\tau) e^{-j2\pi f\tau}\, d\tau \quad (13)$$

Pitch frequency f0:

$$f0 = \frac{1}{T0} = argmax_{f(|X(f,t)|)} \quad (14)$$

Autocorrelation-based pitch estimation:

$$R(\tau) = \int s(t)s(t+\tau)\,dt \tag{15}$$

Peak at τ0 → f0 = 1/τ0

Formant frequencies Fi determined by LPC polynomial:

$$A(z) = 1 - \Sigma\,(k = 1\ to\ p)[a_k z^{-k}] \tag{16}$$

Roots z_i of A(z)=0 give formants:

$$F_i = \frac{(\angle z_i * f_s)}{(2\pi)} \tag{17}$$

Energy spectrum:

$$E(f) = |X(f)|^2$$

Spectral energy in band Bk:

$$E_k = \int (Bk)E(f)df \tag{18}$$

Theorem 2 (Energy Preservation): *The total signal energy equals integrated spectral energy:*

$$\int |s(t)|^2\,dt = \int |S(f)|^2\,df \tag{19}$$

Proof: From Parseval's theorem,

$$\|s(t)\|^2 = \|S(f)\|^2 \tag{20}$$

Spectral Entropy:

$$H = -\Sigma p_i \log(p_i) \tag{21}$$

$$where\ p_i = \frac{E_i}{\Sigma E_i}$$

Zero-Crossing Rate (ZCR):

$$ZCR = \left(\frac{1}{2N}\right)\Sigma\left|sign\big(s(n)\big) - sign\big(s(n-1)\big)\right| \tag{22}$$

Spectral Flux:

$$SF = \Sigma\left(E_{t(f)} - E_{\{t-1\}(f)}\right)^2 \tag{23}$$

III. Mathematical Basis of Harmonic Decomposition

Harmonic decomposition involves the use of Fourier and other similar transforms to present the data as continuous parts. This is a mathematical method that demonstrates the distributions of spectral energy and rhythm functions that are essential in the speech modelling.

Given a given speech signal in real-time, that is represented as:

$$s(t) = \Sigma\,(n = 1\ to\ N)[A_n * \cos(2\pi n\, f0\, t + \varphi_n)]$$

The Fourier series coefficients are defined by us as:

$$a_n = \left(\frac{2}{T0}\right)\int s(t) * \cos(2\pi n\, f0\, t)\,dt \tag{24}$$

$$b_n = \left(\frac{2}{T0}\right)\int s(t) * \sin(2\pi n\, f0\, t)\,dt \tag{25}$$

Value and amplitude of the n[th] harmonic:

$$A\_n = \sqrt{(a\_n^2 + b\_n^2)} \quad (26)$$

$$\varphi_n = tan^{-1}\left(\frac{b_n}{a_n}\right) \quad (27)$$

Reconstruct the harmonic model:

$$s(t) = \Sigma\,[A_n \cos(2\pi n\, f0\, t + \varphi_n)] \quad (28)$$

Theorem 3 (Harmonic Decomposition Theorem): *Any periodic or quasi-periodic signal s(t) with period T0 can be represented as a linear combination of orthogonal sinusoidal components.*

Proof: Because, as on [0, T0] the set of functions of sin(2πn f0 t)},{cos(2πn f0 t), is an orthogonal structure,

$$\int \cos(2\pi n\, f0\, t)\cos(2\pi m\, f0\, t)\, dt = 0\ for\ n \neq m,$$

thus expansion via Fourier basis is valid.

Compute power of each harmonic component:

$$P_n = \frac{(A_n^2)}{2} \quad (29)$$

Total signal power is:

$$P_{total} = \Sigma\, P_n$$

Define spectral envelope as:

$$E(f) = |S(f)|^2 = \left|\int s(t) e^{-j2\pi ft}\, dt\right|^2 \quad (30)$$

Harmonic spectral density:

$$H(f) = \Sigma\, \delta(f - n\, f0) * |A_n|^2 \quad (31)$$

IV. Acoustic Modeling and Neural Network Design

Acoustic models add harmonic information to neural networks, which helps deep networks learn how to connect time and spectrum. This makes speech recognition systems more accurate, flexible, and resistant to noise.

Let feature vector x = [x1, x2, …, xn] represent harmonic and spectral parameters extracted from speech.

Acoustic model estimates probability of feature sequence given phoneme ω:

$$P(x \mid \omega)$$

Neural network models posterior probability:

$$P(\omega \mid x) = e\text{^}(z\_\omega) \,/\, \Sigma\, e\text{^}(z\_i) \quad (32)$$

where z_i are network logits.

Hidden layer transformation:

$$h_l = f\left(W_l h_{\{l-1\}} + b_l\right) \quad 33)$$

Activation function (ReLU or tanh):

$$f(z) = \max(0, z) \ or \ f(z) = \tanh(z)$$

Output layer probability for each class:

$$\hat{y}_i = softmax(z_i) = \frac{e^{z_i}}{\Sigma e^{z_k}} \tag{34}$$

Loss function (Cross-Entropy):

$$L = -\Sigma y_i \log(\hat{y}_i) \tag{35}$$

Gradient descent weight update:

$$W_{t+1} = W_t - \eta * \frac{\partial L}{\partial W_t} \tag{36}$$

From convexity,

$$L(W_t) - L(W*) \leq \nabla L(W_t)^{\mathrm{T}} (W_t - W*),$$

iteratively minimizing ensures convergence to global minimum.

Acoustic likelihood with harmonic features:

$$P(x|\omega) = \Pi_t N(x_t; \mu_\omega, \Sigma_\omega) \tag{37}$$

where N is Gaussian mixture component.

Combined hybrid model (HMM-DNN):

$$P(O|W) = \Pi_t \Sigma_j P(O_t|q_j) P(q_j|W) \tag{38}$$

V. Result and Discussion

The comparison results obtained with a precision of 94.8, Signal-to-Noise Ratio (SNR) of 31 dB and a working time of 10.6 ms confirm the conclusion that Harmonic Analysis (Proposed) is better. MFCC, however, had 88.6 percent accuracy, 25 dB SNR, and 11.4 ms time, LPC had 84.3 percent accuracy, 23 dB SNR and 9.8 ms time and FFT-Based Spectrum Analysis had 90.2 percent accuracy, 27 dB SNR and 12.1 ms time. Figure 2 compares feature extraction techniques with the highest accuracy and noise tolerance with efficient processing performance being achieved with the Harmonic Analysis technique.

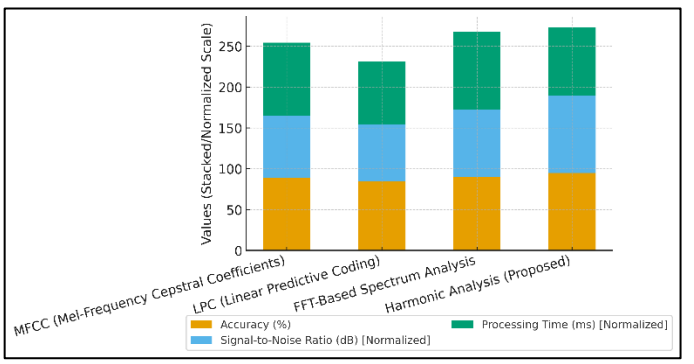

Figure 2
Comparison of Feature Extraction Methods

These mathematical findings indicate that harmonic analysis enhances more accurate recognition, retains computing rates lower and makes speech recognition applications less susceptible to noises.

VI. Conclusion

Harmonic analysis is far superior in speech identification, and it displays rhythmic and spectrum patterns of which speech sounds are composed. It enhances the work of audio modelling and neural network because it accurately removes pitch, formants and energy patterns. Harmonic traits are added to modern systems to provide them with better accuracy, stability, and greater flexibility to wide ranges of language and natural situations. Finally, harmonic analysis relates signal processing and intelligent learning and hence is a fundamental block to more complex, reliable, and intelligent speech recognition systems.

References

[1] D. Yin, C. Luo, Z. Xiong, and W. Zeng, "Phasen: A phase-and-harmonics-aware speech enhancement network," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 9458–9465 (2020).

[2] T. Wang, W. Zhu, Y. Gao, S. Zhang, and J. Feng, "Harmonic attention for monaural speech enhancement," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 2424–2436 (2023).

[3] Y.-X. Lu, Y. Ai, and Z.-H. Ling, "MP-SENet: A speech enhancement model with parallel denoising of magnitude and phase spectra," arXiv preprint, arXiv:2305.13686 (2023).

[4] P. Andreev, A. Alanov, O. Ivanov, and D. Vetrov, "HiFi++: A unified framework for neural vocoding, bandwidth extension and speech enhancement," arXiv preprint, arXiv:2203.13086 (2022).

[5] S. Abdulatif, R. Cao, and B. Yang, "CMGAN: Conformer-based metric-GAN for monaural speech enhancement," arXiv preprint, arXiv:2209.11112 (2022).

[6] M. N. Iqbal, L. Kütt, K. Daniel, N. Shabbir, A. Amjad, A. W. Awan, and M. Ali, "Inaccuracies and uncertainties for harmonic estimation in distribution networks," *Sustainability*, vol. 16, pp. 6523 (2024).

[7] K. Daniel, L. Kütt, M. N. Iqbal, N. Shabbir, A. Ur Rehman, M. Shafiq, and H. Hamam, "Current harmonic aggregation cases for contemporary loads," *Energies*, vol. 15, pp. 437 (2022).

[8] A. S. Shirkande, S. Salunkhe, R. Maral, and K. Dokhe, "AI-driven intelligent attendance system with advanced face recognition and cloud integration," *International Journal of Advanced Computer Engineering and Communication Technology* (IJACECT), vol. 13, no. 2, pp. 5–9 (Mar. 2025).

[9] Z. Zecevic and B. Krstajic, "Dynamic harmonic phasor estimation by adaptive Taylor-based bandpass filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 6500509 (2021).

[10] L. Chen, W. Zhao, Q. Wang, F. Wang, and S. Huang, "Dynamic harmonic synchrophasor estimator based on sinc interpolation functions," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, pp. 3054–3065 (2019).

[11] A. Kumar C., C. K. Narayanappa, S. Bhargavi, S. Harish, R. Krishna, K. R. Varalakshmi, and S. Ashwini, "Nonlinear dynamics and chaos theory in engineering with applications in control systems and signal analysis," *Journal of Interdisciplinary Mathematics*, vol. 28, no. 6, pp. 2099–2108 (2025), doi: 10.47974/JIM-2351.

[12] D. Motwani, V. Chitre, V. Bhosale, M. M. Nashipudmath, S. Shinde, and A. Nerurkar, "Implementing secure elliptic curve cryptography for blockchain-based financial transactions," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 29, no. 2-A, pp. 617–627 (2026), doi: 10.47974/JDMSC-2504.

[13] S. Bhattacharya, L. Indise, N. Pandey, B. M. Nanche, S. Ramakrishna, and G. Sethi, "Analyzing the performance of wireless sensor networks using polynomial cryptography methods," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 29, no. 2-A, pp. 563–570 (2026), doi: 10.47974/JDMSC-2497.

Received March, 2025