Linear Prediction on Cent Scale for Fundamental Frequency (f0) Analysis

Gowriprasad R,[a] **Anand T, R Aravind, and Hema A Murthy**
Indian Institute of Technology Madras, Chennai, India
ee19d702@smail.iitm.ac.in, tanand@cse.iitm.ac.in, aravind@ee.iitm.ac.in, hema@cse.iitm.ac.in

Abstract: Understanding the fundamental frequency and harmonic content of audio signals is crucial for many applications in music analysis, including music transcription, audio synthesis, and genre identification. This study formulates a signal processing approach combining Linear Prediction (LP) analysis and the Cent scale to characterize audio signals' pitch and harmonic structure accurately. Pitch tracking on the LP spectrum in the Cent scale provides more accurate and reliable pitch estimation, especially in the presence of noise or overlapping harmonics. The Cent scale aligns the harmonics of different notes more closely, making it easier to discern the correct pitch.

[Editor: —] https://doi.org/

Date: 3 October 2024

1. Introduction

In audio processing, accurate estimation of the melody contour (the fundamental frequency f0), which refers to identifying the perceived pitch of musical notes in an audio signal, is crucial for music and speech analysis. Being a fundamental auditory characteristic, f0 plays a pivotal role in several key aspects of Music Information Retrieval (MIR) tasks ranging from melodic analysis (Koduri *et al.*, 2012; Ranjani *et al.*, 2011, 2019) and melody extraction (Rao and Rao, 2010; Salamon and Gómez, 2012), music recommendation (Çataltepe and Altinel, 2007), chord recognition (Humphrey *et al.*, 2012) and music transcription (Benetos and Dixon, 2011; Benetos *et al.*, 2018). Accurate f0 information enhances the capabilities of MIR systems, enabling sophisticated analysis (Viraraghavan *et al.*, 2017, 2020).

Due to various applications, different f0 detection methods have been explored, from traditional signal processing to deep learning techniques. Signal processing techniques use time domain characteristics such as the autocorrelation function (Rabiner, 1977; Slaney and Lyon, 1990) and zero crossing rate (Rabiner *et al.*, 1976a), and spectral characteristics (De Cheveigné and Kawahara, 2002; Mauch and Dixon, 2014; Xu *et al.*, 2016). Short-term phase spectrum (Charpentier, 1986), and the sinusoid modeling property of the group delay function were also explored to estimate the f0 (Rajan and Murthy, 2013). The cepstral analysis exploits the properties of f0 harmonics in the spectrum (Ahmadi and Spanias, 1999; Noll, 1964, 1967), while machine learning (Drugman *et al.*, 2018; Liang and Shu, 2022) and deep learning (Kim *et al.*, 2018; Sun *et al.*, 2022; Wei *et al.*, 2023) extracts f0 information from labeled audio and spectral data.

This paper proposes a novel approach by formulating the Linear Prediction (LP) all-pole spectrum modeling on the Cent scale to analyze the f0 and harmonic structure in the context of Indian Art Music (IAM) audio. LP was designed and developed primarily for speech coding (Gerson and Jasiuk, 1990; Spanias, 1994). Further, LP analysis has been used for various applications in the speech community, such as speech enhancement (Yegnanarayana *et al.*, 1999), epoch extraction (Murty and Yegnanarayana, 2008), etc. LP-based features were used in f0 detection of voiced speech (Prasanna and Yegnanarayana, 2004; Rabiner *et al.*, 1976b), and LP-based cepstrum was combined with harmonic product spectrum for f0 detection (Ding *et al.*, 2006). LP analysis was also used to enhance the tracking of partials in music audio processing (Lagrange *et al.*, 2007), onset detection of music instruments (Glover *et al.*, 2011; Gowriprasad and Murty, 2020; Marchi *et al.*, 2014), and adaptive LP analysis was used in musical onset detection for instruments such as violin, piano, etc. (Lee and Kuo, 2006).

[a] Author to whom correspondence should be addressed.

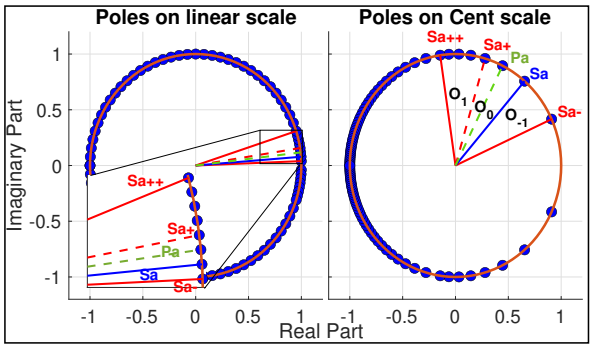

Fig. 1. Z-plane depiction of dilation of frequency in Cent scale.

The autoregressive all-pole model estimates the short-term power spectrum through LP analysis (Makhoul, 1975; Viswanathan and Makhoul, 1975). Linear Prediction analysis computes the envelope of the spectrum by approximating the spectrum by an all-pole model; the roots of the all-pole model correspond to regions of high energy concentration in the signal. In speech, these regions correspond to the resonances of the vocal tract or formants. Perceptual LP is a technique used to perceive speech that follows the Bark scale (Hermansky, 1990). Likewise, we propose another variant of LP formulation on the Cent scale for perceiving music where the perception of music is related to f0, melody, and harmony. The paper is organized as follows. The advantages of the Cent scale for music analysis are outlined to motivate the approach. The Cent scale LP formulation is described and analyzed with various examples. The experimental results are analyzed and discussed in the context of IAM.

2. Why LP on Cent Scale?

LP analysis of spectral envelope gives an equal approximation of the power spectrum $P(e^{j\omega})$ across all frequencies within the analysis band (Hermansky, 1990). This characteristic is at odds with human auditory perception. Human ears perceive pitch in a logarithmic manner, i.e., equal ratios of frequencies are perceived as equal intervals of pitch. This logarithmic perception explains the musical octaves, which signifies a doubling of frequency similar to human auditory perception despite significant frequency differences between the notes. This becomes a primary drawback for all-pole modeling of the harmonics in the music signal when the frequency scale is linear.

The Cent scale, a logarithmic pitch measurement system, offers several advantages over the regular linear frequency scale. One of its primary benefits lies in its ability to represent equal perceptual pitch intervals with equal distances. This means that two points on the Cent scale are separated by the same number of Cents, which sounds equidistant to the human ear. In contrast, equal distances on the linear frequency scale do not correspond to equal perceptual intervals. When using the Cent scale, transposing a musical interval (e.g., a perfect fifth) to a different key becomes a matter of adding or subtracting a fixed number of Cents.

Fig. 1 depicts the poles corresponding to spectral peaks of a hypothetical signal in normal frequency and Cent scales. The scenario considered is sampling freq $fs = 16$KHz, tonic $f_0 = 200$Hz for demonstration purposes. Poles are located at every 100Hz. We can observe in the linear frequency scale that the resolution is poor around the tonic and the necessary octaves. On the Cent scale, the aforementioned limitations are well resolved. The blue solid line corresponds to 0 Cents, the tonic (200 Hz) or Sa in IAM. The green dashed line is the perfect $5th$ (700 Cents), which is 300 Hz, while the dashed red line represents the upper Sa (1200 Cents). The two solid red lines correspond to -1200 and 2400 Cents, enveloping the three octaves $O_1, O_0, 0_{-1}$.

The Cent scale is especially useful in microtonal music, where pitches fall between the standard notes of the chromatic scale. Indian music is rich with microtonal nuances (Agarwal *et al.*, 2013); the tonic is usually decided on stage, and tonal variations often fall between the notes of the Western equal-tempered scale. These microtonal intervals can be precisely defined in Cents. Additionally, the Cent scale facilitates smooth pitch transpositions, as shifting intervals to different keys involve straightforward Cents adjustments. Moreover, human ears are more sensitive to small frequency differences in the middle range of the audible spectrum (Jacobson, 1951). The Cent scale offers a perceptually meaningful, standardized, and versatile way of representing pitch differences, particularly in the middle range of the audible spectrum and all-pole modeling on the Cent scale makes it advantageous in various musical and audio-related applications.

3. Cent Scale LP Formulation

Let $s[n]$ be a windowed time domain signal, and $S(e^{j\omega})$ is its Fourier spectrum. $P(e^{j\omega}) = \mid S(e^{j\omega}) \mid^2$ is the magnitude power spectrum. $P(e^{j\omega})$ is known for $-\pi \leq \omega \leq \pi$, and is real and even, i.e., $P(e^{j\omega}) = P(e^{-j\omega})$. This $P(e^{j\omega})$ is linearly sampled in frequency from $-\frac{fs}{2} \leq f \leq \frac{fs}{2}$ viewed via Discrete Fourier Transform, where fs is the sampling frequency.

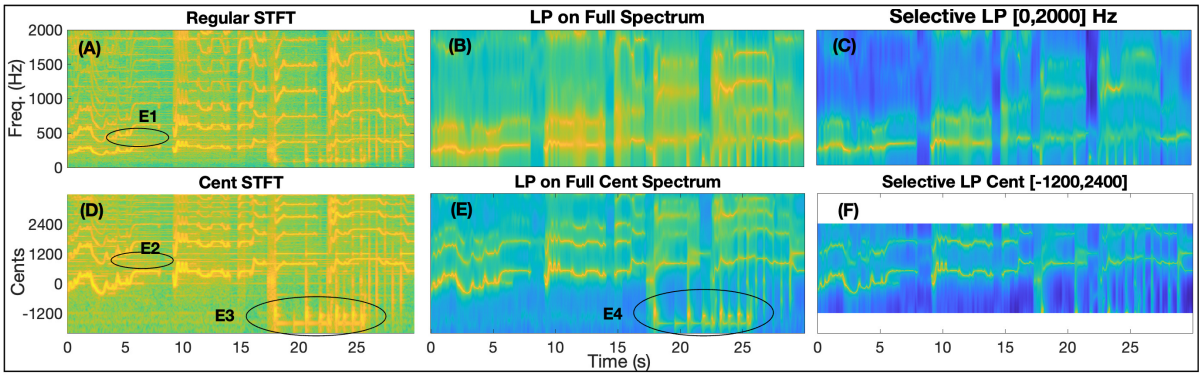

Fig. 2. Spectrograms of a Music Audio in Linear Frequency Scale (A), in Tonic Normalized Cent Scale (D), along with the LP Spectrograms on them (B), (E), and Selective LP Spectrograms (C), (F)

$$\omega_c = 1200 * log_2(\frac{\omega}{\omega_0}) \tag{1}$$

The frequency scale is warped onto the Cent scale using Equation 1 with respect to a tonic ω_0. Let the warped power spectrum be $P(e^{j\omega_c})$ whose samples are spaced log-linearly. The $P(e^{j\omega_c})$ is now resampled linearly in the Cent scale to obtain a linearly spaced power spectrum in the Cent scale $P(e^{j\omega_{c_{lin}}})$. Cubic-spline interpolation is used to replace missing values in the spectrum. Now the task is to fit $P(e^{j\omega_{c_{lin}}})$ in a optimal manner by an all-pole spectrum $\hat{P}(e^{j\omega_{c_{lin}}})$, which is modelled as

$$\hat{P}(e^{j\omega_{c_{lin}}}) = \frac{G^2}{\mid A(e^{j\Omega}) \mid^2} = \frac{G^2}{\mid 1 + \sum_{k=1}^{p} a_k e^{-jk\omega_{c_{lin}}} \mid^2} \tag{2}$$

$A(e^{j\Omega})$ is called the inverse filter, p is the number of poles in the LP model spectrum, and G is a constant gain factor. Now, given $P(e^{j\omega_{c_{lin}}})$ and number of poles p, the task is to determine the parameters a_k, $\forall\ 1 \leq k \leq p$ and G. The Autocorrelation method of all-pole spectral modeling is used to approximate the spectrum $P(e^{j\omega_{c_{lin}}})$. The power spectrum's inverse DFT (IDFT) yields the autocorrelation function dual. The first k + 1 autocorrelation values are used to solve the Yule-Walker equations for the autoregressive coefficients of the k th-order all-pole model. The spectral all-pole modeling details are sufficiently well described (Makhoul, 1975). Algorithm 1 outlines the steps of LP formulation on the Cent scale.

3.1 Selective LP Formulation

The spectral LP formulation can be generalized to fit a selected portion of the spectrum as described in (Makhoul, 1975). The three octaves of the Cent spectrum around the tonic frequency $\omega_c^l = -1200$ to $\omega_c^h = 2400$ Cents are chosen and matched using spectral LP formulation. The selected region is mapped onto the unit circle such that $\omega_c^l \rightarrow 0$ and $\omega_c^h \rightarrow \pi$ (the arrow is read "mapped into"), and the procedure outlined in (Makhoul, 1975) is followed.

ALGORITHM 1: LP formulation on Cent scale: Implementation

1: Transform the audio frame to power spectrum using FFT.
2: Convert the linear frequency axis to the Cent scale using Equation 1.
3: Cubic spline interpolation for linear sampling in Cent scale.
4: Get double-sided power spectrum (DSPS).
5: Compute Inverse FFT of DSPS to get the Autocorrelation function ACF.
6: Compute k of LP coefficients $a_k s$ from the first $k + 1$ number of ACF-values (using Levinson-Durbin recursion).
7: Compute the LP spectrum from G and $a_k s$ using Eqn 2.

3.2 Pitch Estimation

The peaks of the LP spectrum represent the resonances and harmonics. One of the peaks corresponds to the pitch. Estimating the peak locations from a discrete spectrum may compromise resolution, as the number of bins limits the spectral resolution. LP spectrum is an all-pole model where poles model the spectral peaks. LP offers the advantage of computing precise pitch and harmonic values directly from the pole angles of the denominator polynomial $[1, a_k s]$ of Equation 2. Since we have a

ALGORITHM 2: Selective LP formulation and Pitch Detection

1: Select the analysis band interval [ω_c^l to ω_c^h] [-1200 to 2400] three octaves in our case from Step 3, Algorithm 1.
2: Get DSPS for the selected spectrum.
3: Compute $a_k s$ and G as stated in Algorithm 1
4: Compute the pole angles from the LP coefficients – denominator $[1, a_k s]$ polynomial of $\hat{P}(\omega_{c_{lin}})$.
5: Convert back to Cent values from the pole angles using Equation 3 that corresponds to spectral peaks
6: The first peak lying inside the two octaves is considered as pitch.

mapping from the pole angle to the Cent value (Equation 3), we can estimate the Cent value with high precision, eliminating the need for additional peak-picking processes on the LP spectrum. Further, the first pole in the $[-1200, 2400]$ range is considered the pitch. The pitch value in frequency can be obtained via the inverse of Equation 1. Algorithm 2 outlines the steps of selective LP formulation and pitch computation.

$$\omega_{cent} = [(\omega_c^h - \omega_c^l) * (\angle pole/\pi)] - \omega_c^l \tag{3}$$

4. Analysis

Fig. 2 depicts different spectrograms of a music signal. Fig. 2-(A) shows the regular short-time Fourier spectrum. As the signal is sampled at 16KHz, the frequency range is 8KHz, but it is displayed only up to 2000 Hz (in the case of subplots A, B, and C) for better visibility of the pitch and harmonics in the low-frequency region. The harmonics are seen equispaced on the frequency axis, and the desired pitch track also has a low resolution when analyzed in the entire spectrum. The horizontal lines corresponding to the background drone can be observed (denoted by an ellipse E1 and E2).

We can also observe that around 25 seconds, the $3rd$ and $4th$ harmonics have more energy and emphasis than the actual fundamental frequency track. The spectrum also emphasizes the timbre associated with the music, thus making it difficult to detect pitch tracks. Fig. 2-(D) shows the spectrogram in the tonic normalized Cent scale. Here, we can observe that the necessary pitch contour and the octaves have higher resolution. Unlike in the regular frequency scale, the pitch variation distance (vibrato) in the Cent scale is constant across the octaves; this is seen around 10s. We can also observe the background percussion strokes (E3) and the drone (E2).

Fig. 2-(B), and (E) depict the all-pole spectrogram with 30 LP coefficients modeling the entire regular spectrum [0 to 8000 Hz] and Cent STFT [-4580 to 6221 Cents], respectively. All the Cent scale spectrogram plots (i.e., D, E, F) are displayed in the range [-1300 to 2500 Cents]. The all-pole model gives the spectral envelope where the major spectral peaks are modeled with poles, smoothing out the finer spectral characteristics. It is observed that the minor spectral characteristics corresponding to the background drone and the percussion strokes have been smoothed out, enhancing the f0 track. Moreover, the higher-order harmonics are masked while the f0 track is enhanced in the lower frequency region. Nevertheless, the finer details of the f0 track are also masked out in the (B) LP spectrogram derived from the STFT. In (E), the finer f0 details are retained and enhanced, and the percussive strokes are still present (E4).

Fig. 2 (C) and (F) show the selective LP spectrogram on regular STFT and Cent scale, respectively, as described in Section 3.1 with 8 LP coefficients. The analysis band of selective LP in the linear spectrum is from 0 to 2000 Hz, whereas it is from -1200 to 2400 Cents in the Cent spectrum, corresponding to three octaves in music. Even with a lower order of LP (eight), the f0 track is emphasized in the case of selective LP due to limiting the analysis band and having higher resolution.

4.1 Performance with respect to LP order

Fig. 3 depicts the linear spectrum (A, C), tonic normalized Cent spectrum (B, D), and their corresponding LP spectrum with different orders for an audio frame. The linear spectrum is visualized from 0 to 5000 Hz. The Cent spectrum is visualized from -2000 Cents to 5000 Cents, where zero Cents corresponds to tonic frequency. Subplots A and B depict the LP spectrum over the full spectrum on the linear and Cent scales, respectively. The subplots C and D depict the selective LP spectrogram computed on the desired analysis band (0 to 2000 Hz in linear spectrum, -1200 to 2400 Cents in Central spectrum).

The comparison between LP spectra on linear and Cent scales reveals interesting findings. On a linear frequency scale Fig. 3 (A, C), the LP spectrum could not resolve the peaks corresponding to percussion strokes (P1) and f0 (P2), whereas the Cent scale (B, D) showcases these peaks distinctly. Remarkably, the LP spectrum exhibits clear peaks at f0 locations, especially when using the same number of coefficients, offering a detailed representation on the Cent scale. Notably, selective LP spectrum peaks are sharper even with fewer coefficients. Employing just 8 LP coefficients in selective LP (D) produces a spectral fit comparable to 30 to 40 LP coefficients in the entire LP spectrum (B). These observations underscore the efficiency and precision of the Cent scale in capturing essential spectral information, even with reduced computational complexity.

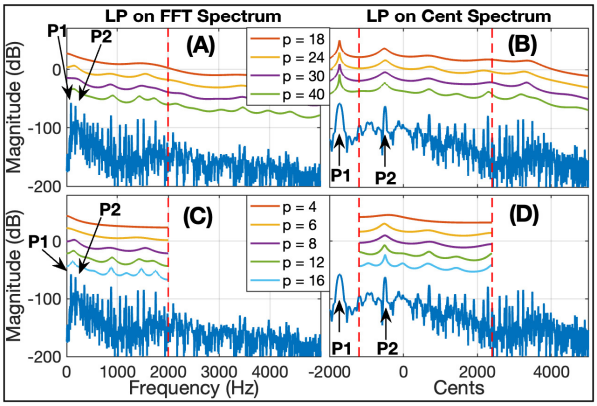

Fig. 3. Comparing LP Spectrum with different orders on linear (A, C) and Cent spectrum (B, D). (A) LP Spectrums on the full range of the linear spectrum. (B) LP Spectrums on the full range of the Cent spectrum. (C) Selective LP Spectrums on the linear spectrum range [0 to 2000Hz]. (D) Selective LP Spectrums on the Cent spectrum range [-1200 to 2400 cents].

Table 1. Performance Scores

Dataset	**Saraga HM**		**Saraga CM**		**Synth CM**	
Metric	RCA	RPA	RCA	RPA	RCA	RPA
Yin	76.6	71.4	75.4	69.1	59.6	59.4
Selec LP-Cent	83.4	74.5	79.6	71.9	67.1	65.7

5. Evaluation

The proposed method is evaluated on the Hindustani (Saraga HM) and Carnatic (Saraga CM) subsets of the IAM dataset (Bozkurt *et al.*, 2018) and SCMS Carnatic music (Synth CM) dataset (Plaja-Roglans *et al.*, 2023). All the audios are sampled to 22500 Hz and used for evaluation. Tonic for the Synth CM dataset is extracted using the methods proposed in (Genís Plaja-Roglans and Thomas Nuttall and Xavier Serra, 2023; Gulati *et al.*, 2014). The f0 detection is performed only on the vocal regions after employing Voice activity detection (VAD). Widely used evaluation metrics such as Raw Chroma Accuracy (RCA) and Raw Pitch Accuracy (RPA) with 50 cent thresholds (Salamon *et al.*, 2014) are used for evaluation. These metrics quantify the percentage of frames where algorithmic outputs match with ground truth within a quarter-tone margin. Evaluation metrics are computed using the reference implementation from mir-eval (Raffel *et al.*, 2014).

Table 1 shows the performance scores of our proposed method and an established signal processing approach for music - Yin (De Cheveigné and Kawahara, 2002). The proposed selective LP uses only 6 LP coefficients, along with the parameters mentioned in Section 3.1, using 100 ms frame length and 10 ms frameshift. We used the Yin approach with the default parameter settings in Librosa library (McFee *et al.*, 2015), which uses a frame length of 2048 samples. The ground truth for Saraga - HM and CM was originally computed using Melodia (Salamon and Gómez, 2012). The results show that the Selective LP method performs better than Yin (De Cheveigné and Kawahara, 2002). The scores are higher in the case of Saraga HM than the Saraga CM and Synth CM, as the CM has more rapid melodic changes and vibrato. Fig. 4 depicts an example of f0 extracted from the proposed selective LP method on an Indian music excerpt using only 6 LP coefficients and with the parameters mentioned in Section 3.1.

Selective LP on a Cent scale is a model-based, unsupervised technique that runs on a single CPU machine with less computational power. The selective LP method is handy for real-time f0 computation. Selective LP worked efficiently even with nominal changes in tonic values, as the change in tonic value essentially shifts the Cent spectrum. With the one-to-one transformation (Equation 1), we can get correct f0 values in Hz even if the peaks are shifted in the Cent scale. The selective LP method can potentially be used as a preprocessing step, enhancing the necessary portion of the melodic contours for further analysis.

6. Conclusion

This paper proposed a novel adaptation of LP all-pole modeling on the Cent spectrum that can be used for enhanced f0 detection. We can achieve greater precision in f0 estimation, especially in scenarios involving musical instruments with continuous f0 glides and microtonal nuances. Experimental results demonstrate the effectiveness of the Cent scale and LP analysis in capturing the f0, even in the presence of spectral complexities. The approximate tonic value is the only requirement for the task. The spectral LP implementation in the python has been shared for research purposes[1].

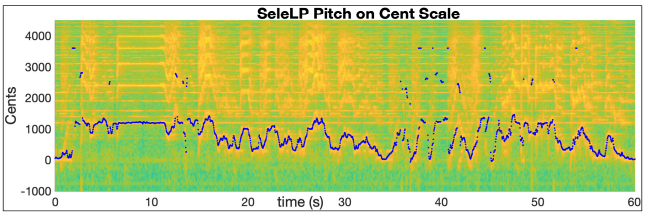

Fig. 4. Fundamental frequency f0 track computed from Selective LP method on Indian vocal excerpt,

References and links

[1] https://github.com/GowriprasadMysore/LP_Cent_Pitch.git

Agarwal, P., Karnick, H., and Raj, B. (**2013**). "A comparative study of indian and western music forms.," in *ISMIR*, pp. 29–34.

Ahmadi, S., and Spanias, A. S. (**1999**). "Cepstrum-based pitch detection using a new statistical v/uv classification algorithm," IEEE Transactions on speech and audio processing **7**(3), 333–338.

Benetos, E., and Dixon, S. (**2011**). "Joint multi-pitch detection using harmonic envelope estimation for polyphonic music transcription," IEEE Journal of Selected Topics in Signal Processing **5**(6), 1111–1123.

Benetos, E., Dixon, S., Duan, Z., and Ewert, S. (**2018**). "Automatic music transcription: An overview," IEEE Signal Processing Magazine **36**(1), 20–30.

Bozkurt, B., Srinivasamurthy, A., Gulati, S., and Serra, X. (**2018**). "Saraga: research datasets of indian art music" .

Çataltepe, Z., and Altinel, B. (**2007**). "Music recommendation based on adaptive feature and user grouping," in *2007 22nd international symposium on computer and information sciences*, IEEE, pp. 1–6.

Charpentier, F. (**1986**). "Pitch detection using the short-term phase spectrum," in *ICASSP'86. IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, Vol. 11, pp. 113–116.

De Cheveigné, A., and Kawahara, H. (**2002**). "Yin, a fundamental frequency estimator for speech and music," The Journal of the Acoustical Society of America **111**(4), 1917–1930.

Ding, H., Qian, B., Li, Y., and Tang, Z. (**2006**). "A method combining lpc-based cepstrum and harmonic product spectrum for pitch detection," in *2006 International Conference on Intelligent Information Hiding and Multimedia*, IEEE, pp. 537–540.

Drugman, T., Huybrechts, G., Klimkov, V., and Moinet, A. (**2018**). "Traditional machine learning for pitch detection," IEEE Signal Processing Letters **25**(11), 1745–1749.

Genís Plaja-Roglans and Thomas Nuttall and Xavier Serra (**2023**). "compiam" https://mtg.github.io/compIAM/.

Gerson, I. A., and Jasiuk, M. A. (**1990**). "Vector sum excited linear prediction (vselp) speech coding at 8 kbps," in *International Conference on Acoustics, Speech, and Signal Processing*, IEEE, pp. 461–464.

Glover, J., Lazzarini, V., and Timoney, J. (**2011**). "Real-time detection of musical onsets with linear prediction and sinusoidal modeling," EURASIP Journal on Advances in Signal Processing **2011**(1), 68.

Gowriprasad, R., and Murty, K. S. R. (**2020**). "Onset detection of tabla strokes using lp analysis," in *2020 International Conference on Signal Processing and Communications (SPCOM)*, IEEE, pp. 1–5.

Gulati, S., Bellur, A., Salamon, J., H.G., R., Ishwar, V., Murthy, H. A., and Serra, X. (**2014**). "Automatic tonic identification in indian art music: Approaches and evaluation," Journal of New Music Research **43**(1), 53–71, https://doi.org/10.1080/09298215.2013.875042, doi: 10.1080/09298215.2013.875042.

Hermansky, H. (**1990**). "Perceptual linear predictive (plp) analysis of speech," the Journal of the Acoustical Society of America **87**(4), 1738–1752.

Humphrey, E. J., Cho, T., and Bello, J. P. (**2012**). "Learning a robust tonnetz-space transform for automatic chord recognition," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 453–456.

Jacobson, H. (**1951**). "Information and the human ear," The Journal of the Acoustical Society of America **23**(4), 463–471.

Kim, J. W., Salamon, J., Li, P., and Bello, J. P. (**2018**). "Crepe: A convolutional representation for pitch estimation," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 161–165.

Koduri, G. K., Gulati, S., Rao, P., and Serra, X. (**2012**). "Rāga recognition based on pitch distribution methods," Journal of New Music Research **41**(4), 337–350.

Lagrange, M., Marchand, S., and Rault, J.-B. (**2007**). "Enhancing the tracking of partials for the sinusoidal modeling of polyphonic sounds," IEEE Transactions on Audio, Speech, and Language Processing **15**(5), 1625–1634.

Lee, W.-C., and Kuo, C.-C. J. (**2006**). "Musical onset detection based on adaptive linear prediction," in *Multimedia and Expo, IEEE International Conference on*, IEEE, pp. 957–960.

Liang, S., and Shu, R. (**2022**). "Extraction of music main melody and multi-pitch estimation method based on support vector machine in big data environment," Journal of Environmental and Public Health **2022**.

Makhoul, J. (**1975**). "Spectral linear prediction: Properties and applications," IEEE Transactions on Acoustics, Speech, and Signal Processing **23**(3), 283–296.

Marchi, E., Ferroni, G., Eyben, F., Gabrielli, L., Squartini, S., and Schuller, B. (**2014**). "Multi-resolution linear prediction based features for audio onset detection with bidirectional lstm neural networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, IEEE, pp. 2164–2168.

Mauch, M., and Dixon, S. (**2014**). "pyin: A fundamental frequency estimator using probabilistic threshold distributions," in *2014 ieee international conference on acoustics, speech and signal processing (icassp)*, IEEE, pp. 659–663.

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., and Nieto, O. (**2015**). "librosa: Audio and music signal analysis in python.," in *SciPy*, pp. 18–24.

Murty, K. S. R., and Yegnanarayana, B. (**2008**). "Epoch extraction from speech signals," IEEE Transactions on Audio, Speech, and Language Processing **16**(8), 1602–1613.

Noll, A. M. (**1964**). "Short-time spectrum and "cepstrum" techniques for vocal-pitch detection," The Journal of the Acoustical Society of America **36**(2), 296–302.

Noll, A. M. (**1967**). "Cepstrum pitch determination," The journal of the acoustical society of America **41**(2), 293–309.

Plaja-Roglans, G., Nuttall, T., Pearson, L., Serra, X., and Miron, M. (**2023**). "Repertoire-specific vocal pitch data generation for improved melodic analysis of carnatic music," Transactions of the International Society for Music Information Retrieval **6**(1), 13–26.

Prasanna, S. M., and Yegnanarayana, B. (**2004**). "Extraction of pitch in adverse conditions," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, Vol. 1, pp. I–109.

Rabiner, L. (**1977**). "On the use of autocorrelation analysis for pitch detection," IEEE transactions on acoustics, speech, and signal processing **25**(1), 24–33.

Rabiner, L., Cheng, M., Rosenberg, A., and McGonegal, C. (**1976**a). "A comparative performance study of several pitch detection algorithms," IEEE Transactions on Acoustics, Speech, and Signal Processing **24**(5), 399–418.

Rabiner, L., Cheng, M., Rosenberg, A., and McGonegal, C. (**1976**b). "A comparative performance study of several pitch detection algorithms," IEEE Transactions on Acoustics, Speech, and Signal Processing **24**(5), 399–418, doi: 10.1109/TASSP.1976.1162846.

Raffel, C., McFee, B., Humphrey, E. J., Salamon, J., Nieto, O., Liang, D., Ellis, D. P., and Raffel, C. C. (**2014**). "Mir_eval: A transparent implementation of common mir metrics.," in *ISMIR*, Vol. 10, p. 2014.

Rajan, R., and Murthy, H. A. (**2013**). "Group delay based melody monopitch extraction from music," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 186–190, doi: 10.1109/ICASSP.2013.6637634.

Ranjani, H., Arthi, S., and Sreenivas, T. (**2011**). "Carnatic music analysis: Shadja, swara identification and raga verification in alapana using stochastic models," in *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, IEEE, pp. 29–32.

Ranjani, H., Srinivasamurthy, A., Paramashivan, D., and Sreenivas, T. V. (**2019**). "A compact pitch and time representation for melodic contours in indian art music," The Journal of the Acoustical Society of America **145**(1), 597–603.

Rao, V., and Rao, P. (**2010**). "Vocal melody extraction in the presence of pitched accompaniment in polyphonic music," IEEE transactions on audio, speech, and language processing **18**(8), 2145–2154.

Salamon, J., and Gómez, E. (**2012**). "Melody extraction from polyphonic music signals using pitch contour characteristics," IEEE transactions on audio, speech, and language processing **20**(6), 1759–1770.

Salamon, J., Gómez, E., Ellis, D. P., and Richard, G. (**2014**). "Melody extraction from polyphonic music signals: Approaches, applications, and challenges," IEEE Signal Processing Magazine **31**(2), 118–134.

Slaney, M., and Lyon, R. F. (**1990**). "A perceptual pitch detector," in *International conference on acoustics, speech, and signal processing*, IEEE, pp. 357–360.

Spanias, A. S. (**1994**). "Speech coding: A tutorial review," Proceedings of the IEEE **82**(10), 1541–1582.

Sun, X., Liang, X., He, Q., Zhu, B., and Ma, Z. (**2022**). "Gio: A timbre-informed approach for pitch tracking in highly noisy environments," in *Proceedings of the 2022 International Conference on Multimedia Retrieval*, pp. 480–488.

Viraraghavan, V. S., Aravind, R., and Murthy, H. A. (**2017**). "A statistical analysis of gamakas in carnatic music.," in *ISMIR*, pp. 243–249.

Viraraghavan, V. S., Pal, A., Aravind, R., and Murthy, H. A. (**2020**). "Data-driven measurement of precision of components of pitch curves in carnatic music," The Journal of the Acoustical Society of America **147**(5), 3657–3666.

Viswanathan, R., and Makhoul, J. (**1975**). "Quantization properties of transmission parameters in linear predictive systems," IEEE Transactions on Acoustics, Speech, and Signal Processing **23**(3), 309–321.

Wei, H., Cao, X., Dan, T., and Chen, Y. (**2023**). "Rmvpe: A robust model for vocal pitch estimation in polyphonic music," arXiv preprint arXiv:2306.15412 .

Xu, X., Zhao, M., Lin, J., and Lei, Y. (**2016**). "Envelope harmonic-to-noise ratio for periodic impulses detection and its application to bearing diagnosis," Measurement **91**, 385–397.

Yegnanarayana, B., Avendano, C., Hermansky, H., and Murthy, P. S. (**1999**). "Speech enhancement using linear prediction residual," Speech communication **28**(1), 25–42.