2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)
31 July - 2 August 2025, Rangpur, Bangladesh

Formant Estimation Using Singular Value Distribution: An Empirical Study

Saikat Chakraborty
Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology
Khulna, Bangladesh
saikatchakraborty4444@gmail.com

Md. Mahbub Hasan
Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology
Khulna, Bangladesh
mahbub01@eee.kuet.ac.bd

***Abstract*— Natural speech signals are often corrupted by noise and exhibit a composite spectral nature. The resonance of the vocal tract shapes the glottal pulses from the vocal fold and produces speech signals. These resonance frequencies, known as formants, are dominant components in the frequency spectrum and typically contain the key information of speech. Each formant is associated to a pair of singular values obtained through singular value decomposition. The frequency bands of lower magnitude in the distribution carry the noise components and irrelevant signals. Keeping pitch and formant components while filtering out other spectral components from speech will reduce noise and irrelevant signals. This article proposes a singular value thresholding technique for formant estimation. The empirical method for estimating the threshold is based on the singular value distributions of synthesized speech signals with varying formant conditions. We applied the proposed technique on a natural Bangla vowel and compared the estimated formant to the LPC-based method. This approach can help easily recognize the formants, which can be used in medical diagnostics, forensics, speech-text applications, and speech synthesis.**

Keywords—Singular Value Decomposition, Formants, Hankel Matrix, Impulse Response

I. INTRODUCTION

Formants are physically defined as poles in a system function expressing the characteristics of a vocal tract. In speech signals, formants are frequency peaks in the spectrum which have a high degree of energy. These are especially prominent in vowels [1]. The speech signals are the output of vocal tract resonance or formant filter where the excitation source is glottal pluses produced by the vocal fold. During speech production, the formant frequency and bandwidth vary dynamically according to the size and shape of the vocal tract. Formant frequency and bandwidth can be used to estimate the phonetic content of speech sounds. Using various linear and spectral analysis methods, formant components are used as speech recognition features [2]. Formant components encapsulate the key linguistic information, and these are pivotal components in speech reconstruction. Speech reconstruction by focusing these components can effectively suppress the noise and irrelevant components of speech and thus improve signal quality [3]. This speech reconstruction process is effectively used in speech compression. The compressed speech signal can be transmitted or preserved effectively through digital modulation techniques with minimum errors. Modern lossy audio codecs (e.g. mp3) implement similar methods where phase, frequency, and amplitude information of significant components are used to create synthetic reproduction of speech [4]. By filtering signals to isolate or enhance specific formants, speech recognition systems can better distinguish between different phonemes and improve accuracy [5]. Polyps in the vocal tract or other abnormalities can also be diagnosed by examining additional formants and their frequencies in the speech signals. Polyps constrict the vocal tract; therefore, it creates a small resonating chamber, and higher resonating frequency. Constriction of the vocal tract can also result in higher or lower formant frequencies [6]. The distribution of the first four formants shows deviation which allows diagnosis of dysphonic patients [7].

Existing methods of formant estimation are based on linear prediction, inverse filter control, Eigen Value Decomposition etc. In the study of Welling and Ney [8], parallel resonator model is used to estimate the formants in speech. This method has a small error rate of 4.2%. Another study describes the process of extracting formant information using eigen value decomposition [9]. The article provides a method to estimate the number of significant formants with the help of eigen value distribution. Using the principal components with the highest values, we can reconstruct the speech signal ignoring the less significant components. The distribution of eigenvalues helps to separate important signal components (formant and pitch) from noises [10].

Singular Value Decomposition (SVD), which is similar to eigenvalue decomposition, can effectively reduce noise and classify the spectrum for speech recognitions. The SVD is an unsupervised and comparatively fast method. In contrast, the recent heat-map approach [11] achieves high accuracy but requires supervised training, while the Bayesian method [12] offers robust statistical error bounds at a higher computational cost and is limited to steady-state vowels. Although researchers are utilizing SVD in noise reduction from speech signals, the criteria of selecting singular components, such as the ones related to formants, are not well documented yet. In this article, we will explore the distribution of singular values (SVs) at different formant conditions to find the reliable patterns of determining the number of formants and frequencies in a noisy speech signal. Based on these patterns, we will empirically estimate the SV threshold for formant estimation of speech signals.

We have organized the article as follows: Section II presents the mathematical formulation of formant filtering, speech signals synthesis, and SVD. Section III explores the singular value distribution with different formant conditions, and formant estimation criteria numerically. Section IV, we investigate the effects of formant component damping on singular value distribution. Section V describes the formant estimation in natural vowel. Sections III, IV, and V collectively describe the results and discussion. Lastly, Section VI summarize the article with important empirical rules for selection formant components of speech signal.

979-8-3315-9694-1/25/$31.00 ©2025 IEEE

II. MATHEMATICAL FORMULATION

For singular value investigation with formant, synthesized speech signals are preferable since their formant parameters can be easily tuned, therefore giving more control over the natural speech signals. The transfer function of a filter having multiple poles can infuse formant characteristics in speech synthesis. The impulse response of the filter describes the dynamic behavior in time domain. A digital filter is characterized by a transfer function, $H(z)$ which can be expressed as:

$$\mathrm{H(z)}=\frac{\mathrm{Y(z)}}{\mathrm{X(z)}} \tag{1}$$

Where, H(z) is the transfer function, Y(z) is the Z-transform of the output signal, and X(z) is the Z-transform of the input signal. We can generate the impulse response function by varying the formant filter characteristics as required.

A. Formants and Impulse Responses

The formant frequency corresponds to the resonant peak of the transfer function of the vocal tract and its width is the formant bandwidth. The bandwidth determines the time-dependent damping ratio of the impulse responses. Using a digital filter, its impulse response can be obtained in the form of sinusoid with specific frequency and desired damping. The parallel representation of vocal tract is a network where multiple formants or resonant filters are parallelly connected. Their individual filter output (convolution of impulse response and glottal pulse) collectively produces speech signals.

B. Association with Poles

The sinusoidal impulse response is produced by a system having transfer function with a pair of poles, usually complex conjugate in nature. The poles that are associated with the formant take position in the Z-plane as shown in Fig. 1. Each formant gives at least one pair of SVs in the distribution which is a significant characteristic of the speech signal [13].

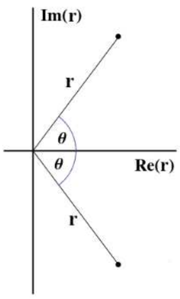


Fig. 1: The complex conjugate poles of the transfer function representing the vocal-tract in Z-plane.

Assume we want to generate an impulse response that has a sampling frequency F_s, formant frequency denoted by f, and bandwidth BW. The radius of the poles associated with the formant is given by:

$$r = 1-\frac{BW}{F_s}\pi \tag{2}$$

And the angle can be written as:

$$\theta = \frac{f}{F_s}\cdot 360° \text{ in degree} \tag{3}$$

$$= \frac{f}{F_s}\cdot 2\pi \text{ in radian} \tag{4}$$

C. Generation of Synthesized Speech

Fig.2 shows the synthesized speech generation process. The output of each formant filter is the convoluted form of impulse response with the excitation, $g(n)$. The composite signal is normalized in the end for simplicity.

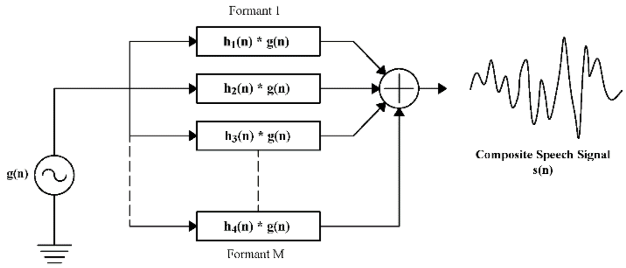


Fig. 2: Synthesis of Speech Signal using parallel resonator model.

The transfer function of the formant filter is given by:

$$\mathrm{H(z)} = \frac{1}{(z - re^{j\theta})(z - re^{-j\theta})} \tag{5}$$

Breaking the expression into partial fraction, we get

$$\mathrm{H(z)} = \frac{1}{2rj\sin\theta}\left[\frac{1}{z - re^{j\theta}} - \frac{1}{z - re^{-j\theta}}\right] \tag{6}$$

Applying inverse Z Transform,

$$h(n) = \frac{1}{2rj\sin\theta}\left[\left(re^{j\theta}\right)^n u(n) - \left(re^{-j\theta}\right)^n u(n)\right] \tag{7}$$

We can normalize the expression by omitting the sine term to prevent change of amplitude with formant frequency.

$$h(n) = \frac{1}{2rj}\left[\left(re^{j\theta}\right)^n u(n) - \left(re^{-j\theta}\right)^n u(n)\right] \tag{8}$$

This is the expression of a sinusoid with the frequency of the formant and has an exponential damping characteristic determined by the bandwidth. The output of the formant is,

$$s(n) = h(n)*g(n) \tag{9}$$

For composite signal generation multiple formants are combined in cascade or parallel. In this study, the parallel configuration is utilized. The synthesized speech signal can be modelled as equation (10).

$$s(n) = \sum_{i=1}^{m} h_i(n) * g_i(n) \tag{10}$$

D. Estimation of singular values

To apply SVD, we need a two dimensional Hankel matrix. Therefore, the speech signal is reorganized in a trajectory structure (Hankel matrix) by pulling the relevant subseries and stacking them as columns. The length of the subseries is called window length, that should be inversely proportional to the lowest frequency of the composite speech signal. The Hankel matrix of a time series exhibits a specific pattern that can be utilized to filter out noise components of a signal [14]. In a Hankel matrix, every column is shifted up by one element from the previous column.

$$A=\begin{bmatrix} s_0 & s_1 & s_2 & \cdots & s_n \\ s_1 & s_2 & s_3 & \cdots & s_{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_m & s_{m+1} & s_{m+2} & \cdots & s_{m+n-1} \end{bmatrix} \quad (11)$$

The basic definition of SVD is given by:

$$\mathbf{A} = \mathbf{USV}^{\mathrm{T}} = \sum_{k=1}^{M} \mathbf{u}_k s_k \mathbf{v}_k^{\mathrm{T}} \quad (12)$$

where, U, S, V are orthogonal matrix containing left singular vectors, diagonal matrix containing SVs and orthogonal matrix containing right singular vectors respectively. Here, T stands for transpose and M is the number of significant singular components. The kth SVD mode consists of a left singular column vector u_k, a singular value s_k and a right singular row vector v_k [15]. Synthesized speech signal generation

For synthesized speech signal, the impulse response of the parallel formant filter has been evaluated according to the equation (8) and shown in Fig. 3 for both damped and undamped conditions. Amplitude variation was kept between -1 to +1. Compared to the undamped, the damped formant impulse response decays exponentially. The synthesized speech signal produced considering two (80Hz and 120Hz) and three (80Hz, 120Hz and 240 Hz) formants according to the equation (10) are shown in Fig. 4. Between the two synthesized speech signals, three formant vocal tract system produces more fluctuating amplitude as the system consists of an additional high frequency (240 Hz) component.

III. Exploring the Singular Value Distribution

For investigation of the distribution of the singular values with different formant numbers, we have synthesized five speech signals. The choice of formant frequencies in synthesizing the speech signals are shown in Table I. For example, the speech signal 1 contains only one formant frequency (F1=80Hz). The speech signal 3 contains 3 formant frequencies (F1= 80Hz, F2= 120 Hz, and F3= 240 Hz). The one-dimensional synthesized speech signal is then converted to a Hankel structure using a window length of 240. After applying SVD on the Hankel matrix, the resultant SVs from the diagonal matrix were normalized to be investigated. The entire process of formant estimation can be summarized in a few steps as shown in Fig. 5. The output of the process is the SV threshold which will be utilized in the estimation of a natural vowel. Fig. 6 shows the distributions for the 5 undamped formant-based speech signals. Since the amplitude of the SVs decrease as the number of formants increases, the corresponding SVs were normalized to fit them all in the same curve. Each signal is represented by a different color in the graph. For example, the blue curve shows the distribution of the first composite signal that contains only one formant. Similarly, the red curve shows distribution for the fourth signal that contains four formant frequencies, 80Hz, 120Hz, 240Hz, and 360Hz, respectively. The first 20 SVs were taken because further values were negligibly small even for the signal with the highest number of formants.

TABLE I
FORMANT FREQUENCIES USED FOR SYNTHESIS OF SPEECH.

Signal sequence	Formant Frequencies (Hz)				
	F1	*F2*	*F3*	*F4*	*F5*
1	80				
2	80	120			
3	80	120	240		
4	80	120	240	360	
5	80	120	240	360	480

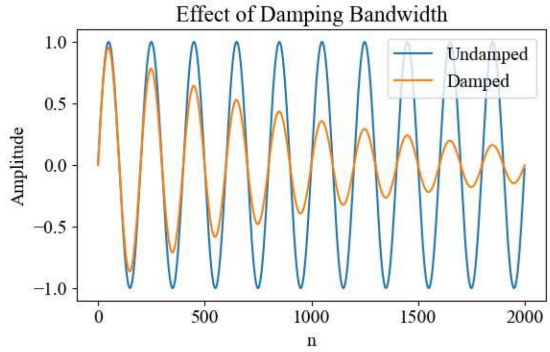


Fig 3: Comparison between damped and undamped impulse response.

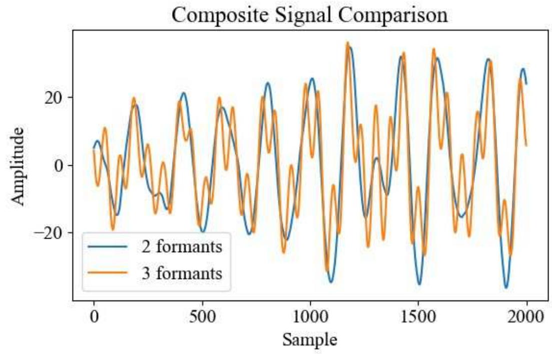


Fig 4: Comparison of synthesized speech signals consisting of two and three formants.

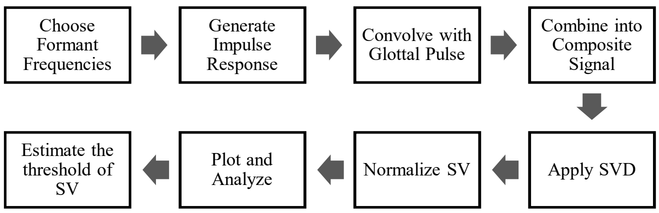


Fig 5: Using SVD to estimate the threshold for the number of formants in speech signal.

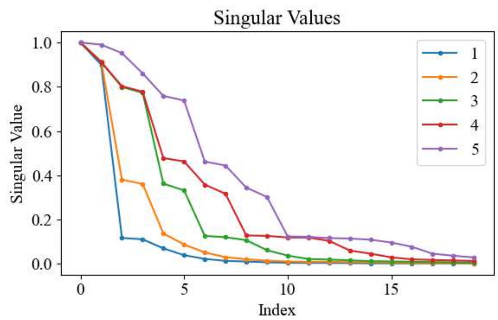


Fig 6: Comparison of singular value distributions for synthesized speech signals with undamped formants. The color legend indicates the number of formants presented in each synthesized speech signals.

TABLE II
NORMALIZED SINGULAR VALUE DISTRIBUTION OF SPEECH SIGNAL CONSIDERING DIFFERENT NUMBER OF FORMANTS.

SVs	No. of formants				
	1	*2*	*3*	*4*	*5*
SV1	1	1	1	1	1
SV2	0.8998	0.9148	0.9118	0.912	0.9907
SV3	0.117	0.3805	0.799	0.8027	0.9523
SV4	**0.1109**	**0.3612**	0.7733	0.7781	0.8609
SV5	0.0697	0.1372	0.3624	0.4778	0.759
SV6	0.039	0.0874	**0.3319**	0.4627	0.7381
SV7	0.0224	0.0513	0.1262	0.3582	0.4621
SV8	0.0132	0.03	0.1204	**0.3165**	0.4442
SV9	0.0105	0.0208	0.1069	0.1282	0.3447
SV10	0.007	0.0146	0.0617	0.1265	**0.3025**
SV11	0.0048	0.0101	0.0371	0.1189	0.1239
SV12	0.0042	0.0092	0.0218	0.1188	0.1216

Table II contains the normalized SVs including the significant ones, obtained from each signal for the undamped condition with a sampling frequency of 16000 Hz. For example, the signal that contains only one formant gives two significant SVs, SV1 and SV2 which are identified by the large difference from the next singular value SV3. For the signal containing two formant frequencies (signal 2), the normalized SVs are 1, 0.9148, 0.3805, 0.3612, 0.1372, ...,0.0092. Now, we need a threshold value, the singular components above this threshold are related to the formants. The singular components having values lower than the threshold are considered irrelevant signals or noises. For developing an empirical rule to approximate the value of threshold of the singular value, we have investigated the numerical values of Table II. The pattern of the singular value distribution suggests that we can approximate the threshold by two steps: first we will select SVs associated with the second largest deviation, then select the higher singular value as the threshold. Then we will count the number of singular components having values equal to or above the threshold. The SVs associated with second largest deviations are (0.1109, 0.0697), (0.3612, 0.1372), (0.3319, 0.1262), (0.3165, 0.1282), and (0.3025, 0.1239) for the speech signal having 1, 2, 3, 4, and 5 formants respectively. So, the threshold SVs are 0.1109, 0.3612, 0.3319, 0.3165, and 0.3025. Table II shows these thresholds in bold. The number of SVs above or equal to the threshold are 4, 4, 6, 8, and 10 for the synthesized speech signals having format 1, 2, 3, 4, and 5, respectively. Since two singular values represent a formant, the numbers of estimated formants are 2, 2, 3, 4, and 5, respectively. But, the actual numbers of formants are 1, 2, 3, 4, and 5 in this case. So, the proposed threshold can determine the number of formants for the signals having two or more formants. But it fails to do so if there is only one formant present. The value of this threshold can depend on many factors such as damping, noise, number of formants etc., all of which reduce the SVs. Nevertheless, the SVs corresponding to non-rhythmic noise stay below the threshold.

The two SVs form a pair, and the pair represents a formant component. Now we need to approximate the maximum deviation between two consecutive SVs, so that we can confirm that these two are in a pair. For this purpose, we have calculated the intra pair distance for the five synthesized speech signals. In Table III, Intra pair deviation (IP), Average SVs (Avg), and allotted frequencies (F) are shown for these speech signals. The second largest deviations of the SVs are 0.0419, 0.224, 0.2057, 0.1883, and 0.1786 for the five synthesized speech signals. Other than for the one formant-based speech signal, these values are higher than corresponding maximum IP deviation, which are 0.1002, 0.0852, 0.0882, 0.088, and 0.0914. So, we can say that the maximum value of intra pair deviation should be lower than or equal to the second largest SV deviation, empirically.

There is a pair of SVs at specific positions in the distribution corresponding to a formant frequency. This position is indicated by the average value of the two SVs in the pair. For example, the frequency 80Hz gives a pair of SVs at around 0.95, 120Hz gives a pair at around 0.35, 240Hz at 0.78, 360Hz at 0.45 and so on.

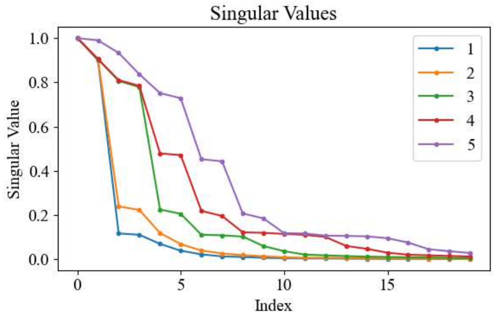
IV. EFFECT OF FORMANT DAMPING

The shape of vocal tract and articulatory muscles determine the formant frequency and bandwidth. In this process, different formant components receive energy differently, due to their distinct damping characteristics. Formant components with different damping can be used to model the natural speech signal. For investigating the effect of damping on the formant components, we have resynthesized the above five speech signals considering the damping factors. We have damped the formants to 5dB level

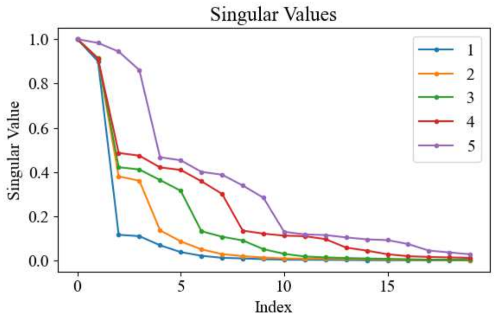
TABLE III
INTRAPAIR DEVIATION, AVERAGE SINGULAR VALUE AND ALLOTTED FREQUENCY OF THE FORMANTS

Pairs	F1			F2			F3			F4			F5		
	IP	*Avg*	*F*	*IP*	*Avg*	*F*	*IP*	*Avg*	*F*	*IP*	*Avg*	*F*	*IP*	*Avg*	*F*
1	0.1002	0.950	80	0.0852	0.957	80	0.0882	0.956	80	0.088	0.956	80	0.0093	0.995	480
2				0.0193	0.371	120	0.0257	0.786	240	0.0246	0.790	240	0.0914	0.907	80
3							0.0305	0.347	120	0.0151	0.470	360	0.0209	0.749	240
4										0.0417	0.337	120	0.0179	0.453	360
5													0.0422	0.324	120

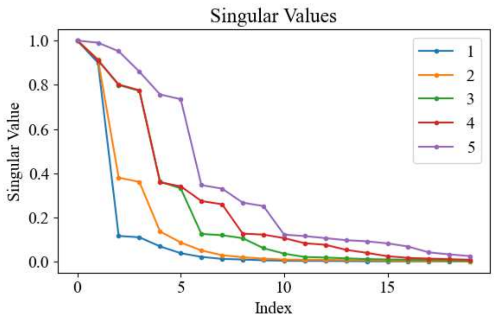
for this purpose. The singular value distribution by utilizing the damped formant frequency at 120 Hz, 240 Hz, 360 Hz, and 480 Hz are evaluated and shown in Fig. 7 (a), (b), (c), and (d) respectively. Like Fig. 6, each figure contains multiple curves, and each of these curves correspond to a signal with a number of formant frequencies denoted in its legend. For example, the red curve shows the distribution for the damped composite signal that contains four formant frequencies.



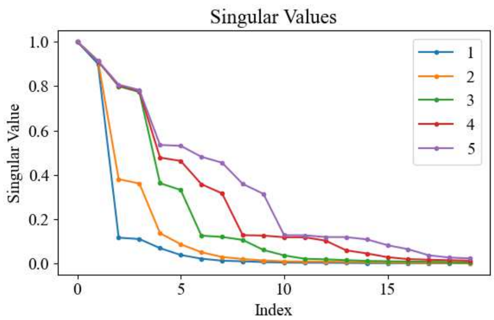
(a) 120 Hz Damped



(b) 240 Hz Damped



(c) 360 Hz Damped



(d) 480 Hz Damped

Fig 7: Singular value distributions with damped formants: (a) 120 Hz, (b) 240 Hz, (c) 360 Hz, and (d) 480 Hz damped; other components as in Fig. 6.

Comparing Fig. 7 to Fig. 6, we can see that for a specific damped formant component, the SVs associated to the formant drops in magnitude, while all the other SVs remain intact. For example, in the distribution of Fig. 7(a), the average value of second singular value pair is around 0.21, but the value was at around 0.35 in undamped condition. The rest of the SVs remain in place as they were in Fig. 6. All the distributions in Fig. 6 exhibits SVs at 0.35 except the first blue line. All the signals had 120Hz frequency in their spectrum except the first one. Therefore, it is concluded that 120Hz causes pairs at around 0.35. Similarly, the SVs at around 0.78 in Fig. 6 drop to 0.41 when 240Hz is damped as shown in Fig. 7(b), SVs at 0.45 drop to 0.26 when 360Hz is damped and shown in Fig. 7(c), and SVs at 0.99 drop to 0.53 when 480 Hz is damped shown in Fig. 7(d). The damping in formant will reduce the amplitude of the SV pair associated with the formant. So, ignoring the low SV components means eliminating the unimportant formants or other rhythmic components irrelevant to the speech signal.

Fig. 7 illustrates that the upper limit of the second largest SV deviation range is the approximated singular value threshold for damped condition. In the undamped condition, any singular value exceeds the threshold will be considered as formant component. But this empirical rule fails to estimate the number of formants for one formant-based speech signal. It should be mentioned that only one formant-based speech signal does not carry perceptible information. So, our proposed empirical threshold range can detect formant in meaningful speech properly.

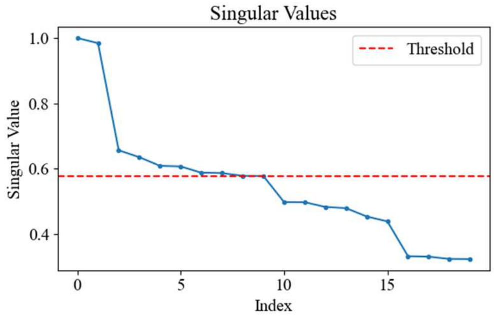
The SV distribution also shows that is that frequency is proportional to the intra pair deviation of the SVs. Lower frequency creates a very large SV deviation, therefore steep slope in the SVs of the pair, whereas around higher frequency the pairs are close together. For example, the 80 Hz frequency causes the highest intra pair gap which is 0.1. Although this gap changes with the number of formants and noise level in the signal, the 120Hz, 240Hz, 360Hz, and 480Hz gives average pair gaps of around 0.04, 0.02, 0.015, and 0.001, respectively.

Based on the pattern discussed above, the number of formants can be estimated by following the procedure outlined below:

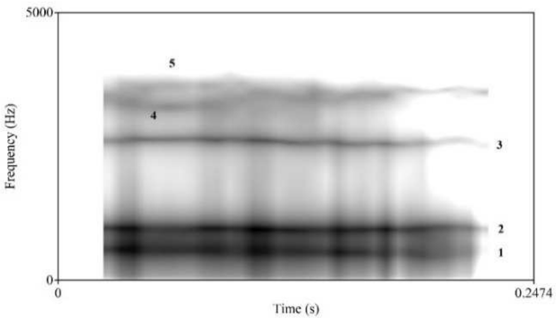
i. *For threshold approximation:* Identify the second largest deviation between the SVs and we can set the upper limit of the deviation as threshold. The SVs from the first singular value to the threshold singular value are associated with the formant components both in undamped and damped formant conditions.
ii. *For SV coupling:* First calculate the deviation between the first and second SV, if the deviation is lower than second largest SV deviation, these two SVs are in a pair and represent a formant. But the deviation is higher than the second largest deviation, first and second SVs are not in pair. Then repeat the procedure for the second and third SV pairs and so on.

V. Formant Thresholding in Natural Vowel

We have applied the SVD on the first Bengali vowel "অ" (phonetically like the English sound "O"), using samples from the Bengali vowel dataset [16]. Fig. 8 (a) shows the SV of the vowel. The upper limit of the second largest deviation is 0.578, which is the estimated threshold value here. There are 10 singular values that are greater than or equal to this threshold, indicating the presence of 5 significant formants. To validate this, we also estimated the formant spectrogram using the LPC method implemented in the Praat software [17], as shown in Figure 8(b). The spectrogram reveals five distinct formant bands. This consistency between the empirical thresholding technique and the LPC-based method supports the reliability of our proposed formant estimation method.



(a) Singular Value Distribution



(b) Formant Spectrogram

Fig 8: SV Distribution and spectrogram for a the first Bengali vowel "অ" .

VI. Conclusion

For formant component estimation in speech signal, SVD is utilized, and this technique can reduce the noise and correlated signals. Rhythmic components in the speech signals show significant pairs of SVs. A significant SV pair represents a formant component in speech. The formant component which has higher value than singular value threshold is considered as significant or dominating formant component. Here, we empirically proposed that the upper bound of the second highest deviation in the singular value distribution is the threshold. For the pairing of the singular components, we proposed that if the deviation of two consecutive singular components is lower than the second largest singular value deviation, then these consecutive singular components from a pair. Here, pairing must start from the first singular component. The intrapair gap between the values within a pair can also be used to estimate the frequency of a formant. Low formant frequency is indicated by high intra pair deviation. With the increase in the number of formants the number of significant pairs increases proportionally but their amplitude decreases. Thus, these findings can help predict the properties of an unknown signal. There are many aspects of signal related science where this operation can have significant impact on, such as detection of communication channels, noise filtering and reconstruction, medical diagnosis and so on.

References

[1] S. Mehrotra, Introduction to EEG- and Speech-Based Emotion Recognition, 1st Edition. 2016.

[2] J. N. Holmes, W. J. Holmes, and P. N. Garner, 'Using formant frequencies in speech recognition', in 5th European Conference on Speech Communication and Technology (Eurospeech 1997), ISCA, Sep. 1997, pp. 2083–2086. doi: 10.21437/Eurospeech.1997-551.

[3] D. H. Klatt, 'Software for a cascade/parallel formant synthesizer', The Journal of the Acoustical Society of America, vol. 67, no. 3, pp. 971–995, Mar. 1980, doi: 10.1121/1.383940.

[4] R. McAulay and T. Quatieri, 'Speech analysis/Synthesis based on a sinusoidal representation', IEEE Trans. Acoust., Speech, Signal Process., vol. 34, no. 4, pp. 744–754, Aug. 1986, doi: 10.1109/TASSP.1986.1164910.

[5] Z. Dong, Q. Ding, W. Zhai, and M. Zhou, 'A Speech Recognition Method Based on Domain-Specific Datasets and Confidence Decision Networks', Sensors, vol. 23, no. 13, p. 6036, Jun. 2023, doi: 10.3390/s23136036.

[6] S. S. Choi and G. H. Zalzal, 'Voice Disorders', in Cummings Otolaryngology - Head and Neck Surgery, Elsevier, 2010, pp. 2876–2883. doi: 10.1016/B978-0-323-05283-2.00204-4.

[7] G. Muhammad, T. A. Mesallam, K. H. Malki, M. Farahat, M. Alsulaiman, and M. Bukhari, 'Formant analysis in dysphonic patients and automatic Arabic digit speech recognition', BioMed Eng OnLine, vol. 10, no. 1, p. 41, May 2011, doi: 10.1186/1475-925X-10-41.

[8] L. Welling and H. Ney, 'Formant estimation for speech recognition', IEEE Trans. Speech Audio Process., vol. 6, no. 1, pp. 36–48, Jan. 1998, doi: 10.1109/89.650308.

[9] A. Islam, Md. R. Hasan, M. Z. Hossain, and Md. M. Hasan, 'The Eigenvalue Distribution of Hankel Matrix: A Tool for Spectral Estimation From Noisy Data', in 2021 24th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh: IEEE, Dec. 2021, pp. 1–6. doi: 10.1109/ICCIT54785.2021.9689845.

[10] H. Hassani, N. Alharbi, and M. Ghodsi, 'Distinguishing chaos from noise: A new approach', International Journal of Energy and Statistics, vol. 2, pp. 137–150, Jun. 2014, doi: 10.1142/S2335680414500100.

[11] Y. Shrem, F. Kreuk, and J. Keshet, "Formant Estimation and Tracking using Probabilistic Heat-Maps", in Interspeech, September 2022, Incheon, Korea.

[12] M. Van Soom, and B. de Boer. "A new approach to the formant measuring problem." Proceedings, vol. 33. no. 1. MDPI, 2019.

[13] S. Sanei and H. Hassani, Singular Spectrum Analysis of Biomedical Signals. Boca Raton: CRC Press, 2015.

[14] M. Ghodsi, N. Alharbi, and H. Hassani, 'The Empirical Distribution of the Singular Values of a Random Hankel Matrix', Fluctuation and Noise Letters, vol. 14, p. 1550027, Sep. 2015, doi: 10.1142/S0219477515500273.

[15] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, 'Singular Value Decomposition and Principal Component Analysis', in A Practical Approach to Microarray Data Analysis, D. P. Berrar, W. Dubitzky, and M. Granzow, Eds., Boston, MA: Springer US, 2003, pp. 91–109, https://doi.org/10.1007/0-306-47815-3_5.

[16] M. R. Hasan and M. M. Hasan, 'Isolated Bengali Vowel and Word Speech Sounds', vol. 1, Sep. 2021, doi: 10.17632/2h6975kdsx.1.

[17] 'Praat: doing Phonetics by Computer'. Accessed: Jun. 22, 2025, https://www.fonuva.nl/praat/.