Journal of Management Engineering and Information Technology (JMEIT)
Volume 6 Issue 1, Feb 2019, Online ISSN:2394 - 8124
Impact Factor: SJIF 6.265 (2016) | 4.564 (2015)
Website: www.jmeit.com | E-mail : editorjmeit@outlook.com | jmeit@outlook.com

ACOUSTICALLY CHARACTERIZING NASAL SOUNDS OF KAMBAATA LANGUAGE

Hussein Seid, *Addis Ababa Science and Technology University, hussien.seid@aastu.edu.et*
Mitiku Tamirat, *Wolkite University, mitikutamirat@gmail.com*

Abstract—this paper analyzes acoustic characteristics of nasal sounds of Kambaata language at their place of articulation by using autoregressive moving average (ARMA) model. The ARMA model is an extension of Linear Predictive Coding (LPC) model, which incorporate the zeros of the transfer function of vocal tract coupled with nasal cavity. Since nasal sound production involves the coupling of nasal cavity, so it can be modeled by ARMA process. Sounds of minimal or nearly minimal pair words, containing singleton or geminated nasal phoneme at the initial, medial or final positions were recorded while read by five male and five female native speakers is collected. Formant and ant-formant frequencies of the collected speech has been extracted and analyzed by using one-way ANOVA test to analyze the differences of nasal sounds. The overall duration measurement is also used to characterize the acoustic nature of nasals. It was observed that Kambaata geminated nasals are found to have statistically longer in overall duration than their singleton conjugates for both females and males in all word positions (initial, medial and final). Anti-formant frequencies are found to have statistically significant difference for both female and males at all positions of the target phoneme. In future, it can be extended for noisy and live environments.

Keywords—Kambaata, nasals, geminated, singleton, acoustic characteristics

I. INTRODUCTION

Nasals are sounds found in all most all known languages of the world [1]. English language has three nasals, all are voiced. Amharic, which is Semitic language, and the official language of Federal democratic republic of Ethiopia [2], has three nasal sounds: /n/, /m/ and /ɳ/ [3]. Kambaata language, the official language of Kambaata people in Kambaata-tembaro zone of south Ethiopia, has two nasal phonemes: /n/ and /m/ [4]. The consonant inventory of Kambaata covers 25 safely established phonemes [4]. Table 1-1 shows the consonant inventory of Kambaata language. In Kambaata, nasals can be spoken in two modes: geminated and singleton. Double letters in contrast to IPA standard is used to represent geminated sounds.

Nasal phonemes are produced when velum is lowered to couple nasal cavity to the oral cavity. In the oral cavity, significant constriction is created by lips and tongue for both /n/ and /m/ respectively [5] [6]. In this case, the outgoing air from the lung have alternative path that is said to be nasal cavity. The phonemes produced in this way are said to be nasal phonemes. This is a point where the nasal sounds differ from the other sounds for the same place and manner of articulation [7].

In this paper, we analyzed the acoustic characteristics of nasals sounds of Kambaata language particularly and produce acoustic phonetic descriptions that characterize them.

II. SPEECH MATERIAL FOR THE STUDY OF NASALS IN KAMBAATA LANGUAGE

Minimal or near-minimal pairs of geminated nasals and their counterpart singleton nasals are selected from Grade One up to Grade Eight textbooks of Kambaata language. The target phonemes in the selected words can be at the initial, medial or final position and can be singleton (at all the three positions) or geminated (at only medial or final position). Tarun and Carol have shown that the position of target nasal phonemes has a significant effect on the characteristics of nasal sounds [8]. Hence, the list of the selected words are classified into five cases: SI (Singleton Initial), SM (Singleton Medial), SF (Singleton Final), GM (Geminated Medial) and GF (geminated final) and five words are selected from the whole text for each /n/ and /m/. Since palato-alveolar, /ɲ/ is very doubtful to be considered as a member of the native Kambaata phone set, hence, this phoneme is not included in this study [4].

Journal of Management Engineering and Information Technology (JMEIT)
Volume 6 Issue 1, Feb 2019, Online ISSN: 2394 - 8124
Impact Factor: SJIF 6.265 (2016) | 4.564 (2015)
Website: www.jmeit.com | E-mail : editorjmeit@outlook.com | jmeit@outlook.com

Table I. Consonant inventory of Kambaata [4]

	PLACE OF ARTICULATION → / MODE OF ARTICULATION ↓			labial	alveolar	palato-alveolar	velar	glottal
Obstruents	Stops		voiceless		t	t͡ʃ	k	ʔ
			voiced	b	d	d͡ʒ	g	
			glottalic	p’	t’	t͡ʃ’	k’	
	Fricatives		voiceless	f	s	ʃ		h
			voiced		z	(ʒ)		
Sonorants	Nasals			m	n			
	Liquids	Vibrants	plain		r			
			glottalized		r’			
		Laterals	plain		l			
			glottalized		l’			
Glides				w		y³		

Each identified word, embedded in an acceptable legal carrier sentence, is recorded while it is read by native speakers (five males and five females of middle aged) of the language five times separately with a brief pause after each utterance. It was ensured that the speaker produced the sentences in natural constant tone and speaking style. The recording is carried out in a relatively noise free videoconference room using an ordinary head-mounted Solic microphone (SLR-840MV) with a Sonny (ICD-UX533) digital flash voice recorder. For transcription of the speech file, the *WaveSurfer* tool was used with default settings. All other analysis was carried out using Matlab R2013a version.

For analysis of nasals, seven parameters are used: overall duration, the first three formants and the first three ant-formant frequencies. The overall duration is measured as the time interval from start of target nasal phoneme to the end of target nasal phoneme. The formants are most basic acoustic speech parameters and often measured as amplitude peaks in the frequency spectrum of the sound wave [9] [10]. Formant frequencies are very important in the analysis of the nasal consonants [11]. Ant-formant is the reduction of energy at some particular point of frequency, which is expected for the nasal murmur. These ant-formant frequencies are used to characterize an appreciable amount of coupling of the nasal cavity with oral cavity [12]. These seven measurements were done for singleton and geminated nasal phonemes and the mean (μ) and standard deviation (σ) of each measurement for each target sound are obtained. Sample data of this measurement (measurements of singleton /n/ at initial position for all speakers) is given in Table 4-1. The statistical significance of the difference between geminated and singleton nasals, among target phoneme position in a word and between the two genders is checked by using the one-way ANOVA test for the significance level of 95%.

III. MEASUREMENTS OF FORMANTS, ANTI-FORMANTS AND OVERALL DURATION

This section specifically describes the method that is used to extract required features from speech wave files. As stated earlier, seven features are extracted in this study by using autoregressive moving average (ARMA) technique. The block diagram of the procedure is shown in Figure 3.1.

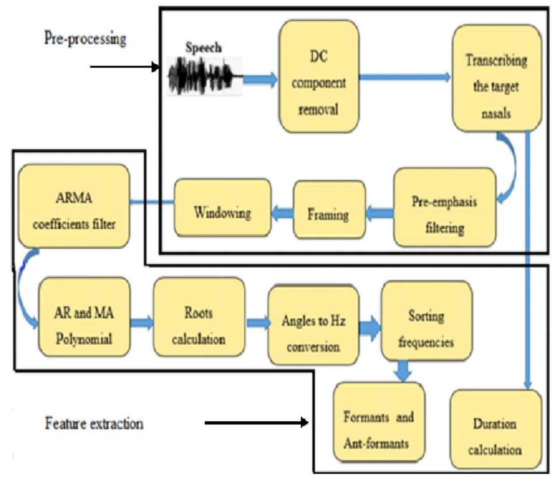


Figure II. Block diagram of formant and anti-formant frequencies extraction

The Linear predictive coding technique (LPC) was used for estimation of the formant frequencies. The basic idea behind this technique is that a speech sample can be approximated as linear combination of previous speech samples [13] [14]. Formants can be then computed from power spectral density. The leading method of formants approximation is based on source-filter separation method i.e. coefficients of LPC are inferred from the windowed speech frames [15].

The LPC estimation procedure is for all-pole systems, not for pole-zero systems. Since during the production process of nasal sounds, the coupling of oral and nasal cavity introduce the zeros at some point of frequency [12], nasal production system is approximated by a pole-zero system. Hence, an ARMA procedure, which is an

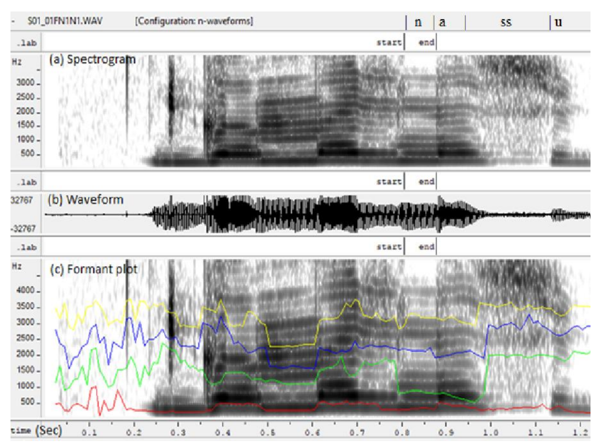


Figure I. (a) Spectrogram, (b) waveform, and (c) formant plot of the carrier sentence containing the word /nassu/.

Journal of Management Engineering and Information Technology (JMEIT)
Volume 6 Issue 1, Feb 2019, Online ISSN:2394 - 8124
Impact Factor: SJIF 6.265 (2016) | 4.564 (2015)
Website: www.jmeit.com | E-mail : editorjmeit@outlook.com | jmeit@outlook.com

Report								
Groups		OD (msec)	F1	F2	F3	F'1	F'2	F'3
initial singleton /n/ uttered by female	μ	83.16	245.04	674.05	2014.40	1073.63	1725.55	2635.14
	σ	23.59	17.82	224.06	266.54	159.06	205.99	242.45
initial singleton /n/ uttered by male	μ	90.56	249.40	1171.82	2339.72	831.74	1579.79	2376.65
	σ	17.39	37.34	397.86	263.62	173.98	226.05	302.62

Table II. Mean and variance values of overall duration (OD in ms), formant frequencies (F1, F2 and F3 in Hz) and anti-formant frequencies (F'1, F'2 and F'3 in Hz) measurements of singleton /n/ at initial position for all speakers

extension of LPC to model pole-zero systems is used in this research to extract formant and ant-formant frequencies.

The process of nasal production is assumed to be an ARMA process of order (p,q) which can then be characterized by a linear difference equation of the form.

$$\sum_{k=0}^{p}(a_k y[n-k]) = \sum_{l=0}^{q}(b_k x[n-l])$$

$$\Rightarrow y[n] = -\sum_{k=1}^{p}(a_k y[n-k]) + \sum_{l=0}^{q}(b_k x[n-l]) \ldots\ldots\ldots\ldots\ldots\ldots 3.1$$

Where x[n] is input, y[n] is output signal, a_k and b_k are coefficients of output and input signals respectively, p and q are order of output and input signals respectively and t denotes time iteration [12]. The transfer function of the process is then

$$H(z) = \frac{B_q(z)}{A_p(z)} = \frac{b_0+b_1z^{-1}+b_2z^{-2}+\cdots b_qz^{-q}}{a_0+a_1z^{-1}+a_2z^{-2}+\cdots a_pz^{-p}} \quad \ldots\ldots\ldots 3.2$$

The coefficients b_j and a_i are referred to as moving average (MA) and autoregressive (AR) parameters respectively, with a_0=1 and b_0=1.

However, the approximation of formants and anti-formants from that of the spectral shape of vocal tract property is nontrivial as not all peaks or bottoms in the spectrum are caused by vocal tract resonances or anti-resonance. Therefore, in this study, the ARMA coefficients are treated as AR polynomial and MA polynomial in order to catch the poles and zeros of the linear differences filter respectively. Their roots are computed and then only positive frequencies up to half of the sampling frequency are taken for computation and for the estimation of formant or anti-formant frequencies. The order of AR and MA, denoted by p and q respectively, is computed using Equation 3.3, where fs is the sampling frequency. This is a classical way used as the thumb's rule for estimating formants apart from other methods [16].

p=q=2+fs/1000......................................3.3

Hence, in this research p and q are calculated to be 10.

Estimation of Formant and Anti-formant Frequencies by ARMA Root-Solving Procedure

For a given a speech signal, each frame of speech to be analyzed is denoted by the N-length sequence of x[n]. Then the ARMA coefficients of each windowed speech frame are calculated using ARMAX function of Matlab. Initial estimates of the formants and ant-formants are computed by solving the complex roots of the AR and MA polynomials respectively [12]. The procedure assured that all the possible formant and anti-formant candidate frequencies are obtained.

Formant and Anti-Formant Extraction Algorithm

The formant and anti-formant frequencies of the target phonemes are obtained by using the following algorithm. The algorithm calculates the frame frequencies.

ALGORITHM Extract Frequencies given a speech signal and a transcription file

Step 1. Remove a DC bias of the given signal using the Equation 3.4.

$$s[n] = s[n] - \frac{1}{L}\sum_{n=1}^{L}(s[n]) \ldots\ldots\ldots\ldots 3.4$$

Where s[n] is the speech signal and L is the length of the signal.

Step 2. Using the information in the transcription file, segment the target region of the signal.

Step 3. Segment the target region into 20 msec frame with 10 msec overlap.

Step 4. calculate the degree of AR and MA using the Equation 3.3

Step 5. For each frame

5.1. Calculate AR and MA coefficients using the ARMAX (x, [p, q]) function of Matlab.

5.2. Calculate the roots of AR and AM coefficients by using ROOTS (POLYDATA (coefficients) function of Matlab and select roots only with positive angle.

5.3. Calculate the formant and anti-formant frequencies from the roots of AR and MA

Journal of Management Engineering and Information Technology (JMEIT)
Volume 6 Issue 1, Feb 2019, Online ISSN: 2394 - 8124
Impact Factor: SJIF 6.265 (2016) | 4.564 (2015)
Website: www.jmeit.com | E-mail : editorjmeit@outlook.com | jmeit@outlook.com

obtained in step 5.2 respectively. Equation 3.5 is used to calculate the frequencies.

5.4. $F_i = tan^{-1}\frac{imaginary(z)}{real(z)} \times \frac{f_s}{2\pi}$ ………………3.5

5.5. Sort the obtained formant and ant-formant frequencies so that the first, the second and the third frequencies is taken as the first, the second and the third formant and anti-formant frequency respectively.

5.6. Repeat Step 5.1-5.4 for all frames

5.7. Calculate the mean of formant and anti-formant of all frames and return the mean values.

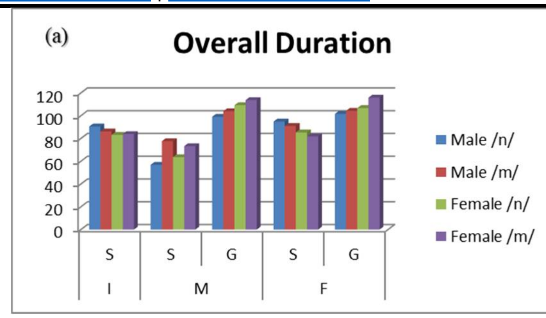
5.8. Overall Duration of Nasal Phonemes

5.9. The duration of the consonant is measured from the onset of sonorant phonation to the constriction release [8]. Analysis of nasal phonemes overall duration was conducted on target phonemes in the speech corpus to differentiate velar nasals from bilabial nasals with their position in a given word and germination and singleton.

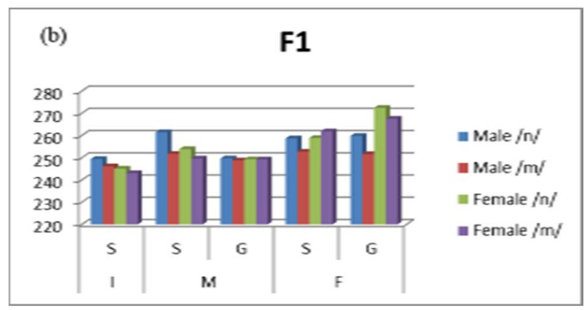
IV. RESULTS AND DISCUSSION

The mean and the variance of overall duration, the first three-formant frequencies and the first three anti-formant frequency measurements for both genders at the same word position in the same phonation style of the two nasals of the Kambaata language is summarized in Table 4.1. To test the variations of the extracted features of the target nasal phonemes, a one way **ANOVA** analysis is conducted to the confidence level of 95%. It is found that initial and final singleton /n/ uttered by females has less overall duration than that of uttered by males. However, no statistically significant difference is found for the initial singleton /m/. As can be seen in Figure 4.1(a), the overall duration of both /n/ and /m/ is almost the same at the initial and medial singleton scenarios. The result shows that there is no statistically significant between these two phonemes in singleton initial position for both males and females.

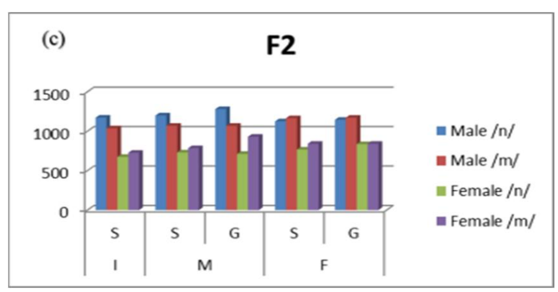
Since Kambaata has both singleton middle and geminated middle nasals, these two groups of sounds were analyzed and the result shows that for both singleton middle velar and bilabial nasals uttered by males have long mean overall duration than for the same sounds uttered by females. For geminated nasal sounds in middle position of a word, there is no significant difference between /n/ and /m/ uttered by both males and female. However, the mean overall duration of geminated middle velar sounds uttered by males is longer than for the same sounds uttered by females.



As can be seen in Figure 4.1(a), there is no significant differences in overall duration for the final singleton and geminated /n/ and /m/ uttered by both females and males. However, the overall duration of final singleton /n/ uttered by males is less than that of uttered by females? Whereas final geminated /m/ uttered by female has less overall duration than their counterpart-geminated sounds uttered by females. In general, overall duration measurement significantly discriminates middle singleton /n/ from middle singleton /m/.



As can be seen in Figure 4.1(b), it is found that both phonemes found in word initial singleton, middle singleton, medial geminated and final singleton have similar first formant frequencies. However, slight difference is observed for final geminated nasals. The first formant frequencies of final geminated /n/ uttered by females is greater than that of final geminated /n/ and geminated /m/ uttered by males.



Journal of Management Engineering and Information Technology (JMEIT)
Volume 6 Issue 1, Feb 2019, Online ISSN: 2394 - 8124
Impact Factor: SJIF 6.265 (2016) | 4.564 (2015)
Website: www.jmeit.com | E-mail : editorjmeit@outlook.com | jmeit@outlook.com

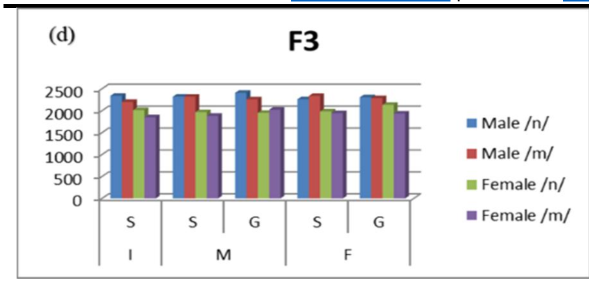
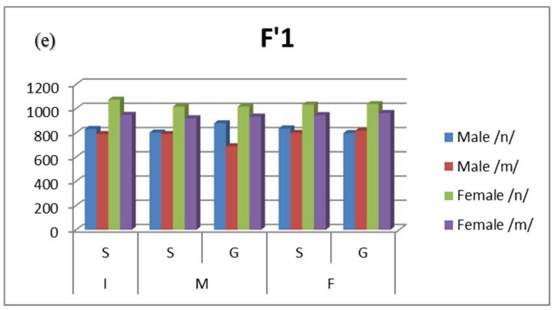
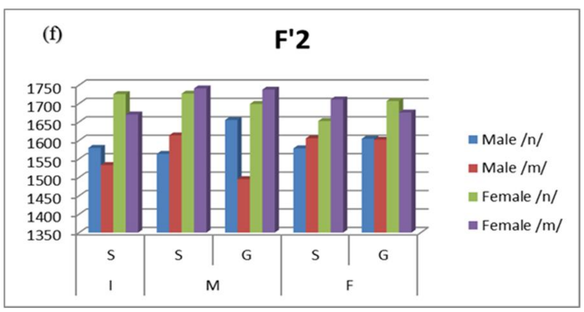


Figure 4.1(c) and (d) show F2 and F3 can significantly separate female speakers from that of male speaker. Male speakers seem to have higher second and formant for both /m/ and /n/. However, both target phonemes demonstrated similar F2 and F3 in similar scenarios.





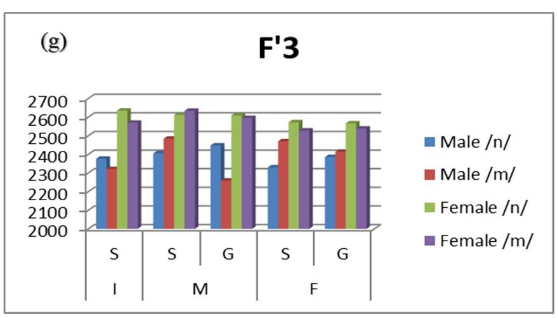


Figure III. 4.1 Summary of (a) overall duration measurement (b) F1 measurement. (c) F2 measurement (d) F3 measurement (e) F'1 measurement. (f) F'2 measurement (g) F'3 measurements

On another hand, as can be seen from Figure 4.1(e), (f) and (g), female speakers are found to have higher anti-formants than that of male speakers. Consistent F'1 is also observed for both /m/ and /n/ across all scenarios.

V. SUMMARY OF FINDINGS

Kambaata is one of the languages to have nasal sounds among the Ethiopian languages. It has velar and bilabial nasals which can be spoken in geminated and singleton phonation. In this research study the overall duration, the first three formant frequencies and the first three anti-formant locations were measured for both nasal sounds with their corresponding word position for both female and male speakers. Both geminated and singleton phonemes were included in the study to characterize them in terms of the list parameters.

The results indicated that the first ant-formant of /n/ is to converge to the frequency range of 798-878 Hz and 1016-1073 Hz for male and female speakers respectively. Similarly, /m/ is found to have first anti-formant frequency range of 689-817 Hz and 920-964 Hz for male and female speakers respectively.

The singleton nasal sounds have significantly less overall duration than their geminated nasal conjugates. Geminated nasal sounds have longer overall duration at the word initial, middle and final position than singleton /m/ and /n/ uttered by both female and male speakers. The zeros associated with anti-resonances are found to be significant variations for word position, gemination/singleton and gender of the speakers. However, for formant frequencies, it is found that there is no statistically significant difference seen for our data.

CONCLUSION

Since conventional LPC model cannot efficiently model source/filter characteristics of the acoustics of nasal sound production mechanism, ARMA model, which is extension to LPC analysis, is used to estimate zeros that are associated with anti-resonances. The results indicated that the first ant-formant of /n/ is to converge to the frequency range of 798-878 Hz and 1016- 1073 Hz for male and female speakers respectively. Similarly, /m/ is found to have first anti-formant frequency range of 689-817 Hz and 920-964 Hz for male and female speakers respectively.

Generally, Formant and Antiformant frequencies can be used to separate /m/ from /n/ in different word positions. Duration feature is also found to be good feature to identify the phonation style (geminated or singleton) of the nasal sounds. The result is very promising to uniquely characterize nasal sounds specifically nasals of Kambaata.

RECOMMENDATIONS

The research work reported in this study is limited to relatively noise free, single channel, ten speakers recorded speech corpus and this study may be extended for noisy multichannel and live data. The new phonetic data obtained from this study may be explored for incorporating in speech systems of the language to improve the performance of the systems. The results of the research reported in this thesis encourage conducting similar types of investigation for other classes of sounds in Kambaata language. The combined results of such investigations can ultimately lead to design of full-fledged automatic segmentation and labeling systems or at least to improve the output of semi-automatic segmentation and labeling systems.

Journal of Management Engineering and Information Technology (JMEIT)
Volume 6 Issue 1, Feb 2019, Online ISSN: 2394 - 8124
Impact Factor: SJIF 6.265 (2016) | 4.564 (2015)
Website: www.jmeit.com | E-mail : editorjmeit@outlook.com | jmeit@outlook.com

From the view of the advancement of speech technology, Kambaata is one of the minority languages of the world. This research work may be considered as one of the few initial efforts in producing important data for the development of quality speech system. Therefore comprehensive and systematic study of phonetic structure of the language is required to improve the quality of speech synthesis, speech recognition, speaker identification, gender identification and other systems in Kambaata language.

REFERENCES

[1] D. Crystal, The Cambridge encyclopedia of the English language, Ernst Klett Sprachen, 2004.

[2] H. S. Worku, S. Rajendran and B. Yegnanarayana, "Acoustic characteristics of ejectives in amharic," *Tenth Annual Conference of the International Speech Communication Association,* pp. 2287--2290, 2009.

[3] I. P. Association, Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet, Cambridge University Press, 1999.

[4] Y. Treis, "A grammar of Kambaata," *Part 1: Phonology, Morphology, and Non-verbal Predication,* 2008.

[5] K. Williamson, "More on nasals and nasalization in Kwa," *Studies in African Linguistics,* vol. 4, no. 2, p. 115, 1973.

[6] R. Blust, "Nasals and nasalization in Borneo," *Oceanic Linguistics,* pp. 149--179, 1997.

[7] T. F. Quatieri, "Production and Classification," in *Discrete-time speech signal processing: principles and practice,* New Delhi, Pearson Education India, 2006.

[8] T. Pruthi and C. Y. Espy-Wilson, "Acoustic parameters for automatic detection of nasal manner," *Speech Communication,* vol. 43, no. 3, pp. 225--239, 2004.

[9] K. Amino, T. Sugawara and T. Arai, "Idiosyncrasy of nasal sounds in human speaker identification and their acoustic properties," *Acoustical science and technology,* vol. 27, no. 4, pp. 233--235, 2006.

[10] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," *Proceedings of the institute of phonetic sciences,* vol. 17, no. 1193, pp. 97--110, 1993.

[11] O. Fujimura, "Analysis of nasal consonants," *The Journal of the Acoustical Society of America,* vol. 34, no. 12, pp. 1865--1875, 1962.

[12] N. Ouaaline and L. Radouane, "Pole--zero estimation of speech signal based on zero-tracking algorithm," *International journal of adaptive control and signal processing,* vol. 12, no. 1, pp. 1--12, 1998.

[13] J. D. Markel and A. J. Gray, Linear prediction of speech, vol. 12, Springer Science & Business Media, 2013.

[14] U. Shrawankar and V. M. Thakare, "Techniques for feature extraction in speech recognition system: A comparative study," *arXiv preprint arXiv:1305.1145,* 2013.

[15] S. Pearson, "Extracting formant-based source-filter data for coding and synthesis employing cost function and inverse filtering". US Patent US Patent 6,195,632, 27 February 2001.

[16] S. McCandless, "An algorithm for automatic formant extraction using linear prediction spectra," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 22, no. 2, pp. 135--141, 1974.