FREQUENCY WARPED ALL-POLE MODELING OF VOWEL SPECTRA: DEPENDENCE ON VOICE AND VOWEL QUALITY

Pushkar Patwardhan and Preeti Rao

Department of Electrical Engineering
Indian Institute of Technology, Bombay, India 400076
pushkar@ee.iitb.ac.in, prao@ee.iitb.ac.in

ABSTRACT

We address the problem of compactly representing the discrete spectral amplitudes of vowel sounds produced by a sinusoidal model. A study of frequency warped all pole model representation of spectral amplitudes has been presented. It has been generally accepted that incorporating Bark scale frequency warping in the all-pole modeling improves the perceived accuracy of the modeled sound. However our study suggests that whether such frequency warped all-pole modeling would improve the modeling accuracy depends on the nature of the vowel as well as the voice. We propose an alternative warping function which may be used to improve the modeling accuracy more universally.

1. INTRODUCTION

The problem of representing the envelope of a discrete spectrum is common to many applications in sound synthesis and coding wherever periodic signals arise. An example is the coding or synthesis of voiced speech based on sinusoidal models [1] in which the parameters are the pitch and harmonic spectral amplitudes. The number of discrete spectral amplitudes depends on the number of harmonics within the frequency bandwidth and therefore on the fundamental frequency. In the context of low bit rate speech coding, this set of spectral amplitudes must be modeled and quantized as compactly as possibly with minimal loss in perceptual accuracy.

Over the years various techniques have been proposed to represent the variable number of discrete spectral amplitudes. Non-parametric methods such as scalar quantization and variable dimension vector quantization [2] and DCT have been applied to quantize spectral amplitudes. However, at very low bit rates parametric methods such as linear predictive (LP) modeling are far more efficient [1]. A set of linear prediction coefficients approximate the spectral amplitudes by another set of spectral amplitudes which are samples of the all-pole modeled spectral envelope at the harmonic frequencies. While approximating the actual envelope by an all-pole envelope, LP modeling minimizes the integrated ratio over frequency of actual spectrum to the approximate spectrum [3]. One of the properties of such a cost function is better modeling at the peaks than the valleys of the spectrum. Such a property forces the spectral envelope to model the pitch harmonics rather than formants for high pitched speakers resulting in underestimated formant bandwidths [4]. There have been several approaches to overcome this problem. These include the envelope interpolated LP proposed by Hermansky [4], in which a smooth envelope is first fitted to the discrete amplitudes via interpolation and then all-pole modeling is carried out of this smooth envelope. Other approaches such as discrete all-pole modeling (DAP) [5] and COSH distance [6] are based on modifying the cost function that is minimized during modeling to consider modeling errors only at the harmonics. Also widely used is the use of predistortion of spectral envelope so that the final error after usual LP modeling (and then restoring the spectral amplitudes by the inverse mapping) is perceptually more acceptable [7]. This has the advantage over cost-function modification that the all-pole modeling itself can be implemented using available efficient techniques.

In this work, we consider the all-pole modeling of narrow-band speech vowel spectra by envelope -interpolated LP [4]. Further, frequency-scale warping, a popular method to improve the perceptual accuracy of the model fit at low model orders, is investigated for sounds of different voice and vowel qualities. In the following sections, we present an introduction to all-pole modeling of spectral amplitudes and to frequency warped all-pole modeling. We investigate the influence of speaker and vowel quality on the perceived modeling error by subjective experiments. An objective distance

This work has been supported in part by a research grant from D.R.D.O., Govt. of India.

measure based on an auditory model is used to obtain an insight into the experimental results.

2. ALL POLE MODELING OF DISCRETE SPECTRA

A typical frequency domain all-pole modeling process [4] has been illustrated in Figure 1. The harmonic spectral amplitudes obtained from a frequency domain analysis of the signal are interpolated to a fixed frequency spacing obtain a smooth spectral envelope which passes through the estimated spectral amplitudes. The power spectrum is computed from the interpolated envelope which is followed by computation of autocorrelation function via IDFT. The Levinson-Durbin algorithm is applied to obtain the all-pole coefficients, which represent the all-pole envelope. The spectral amplitudes are recovered by sampling the spectral envelope represented by the all-pole coefficients at the harmonic frequency locations. It has been reported in [1] that for narrow-band speech vowels, an all-pole model order in the range 16 to 22 is typically required for adequate perceived quality. In the context of speech coding, it is of interest to use as low a model order as possible to minimize the bit allocation.

The harmonic spectral amplitudes estimated from the analysis of an input sound can be looked upon as samples of the underlying source excitation-vocal tract frequency response corresponding to the sound. Since the spectral envelope used for the LP modeling is obtained by smooth interpolation between the harmonic amplitudes, it is expected that it will reflect the spectral details of the actual underlying source-tract envelope only for low-pitched sounds. For high-pitched voices the underlying spectrum is only sparsely sampled and therefore is typically smoother due to the larger extent of interpolation. This is expected to impact the order of the LP model required to achieve a good spectral fit. It is observed that female speech is modeled better than male speech at a given LP model order. The superior quality of synthesized speech for female voices at a given all-pole model order is attributed to the relatively small number of harmonic amplitudes to be approximated as well as to the higher smoothness of the interpolated spectral envelope at higher fundamental frequencies [8]. Thus we see that the model order used in a coding framework will be largely driven by the requirements of perceived quality for low pitched voices. In the next section, we investigate the application of perceptual warping of the frequency scale before all-pole modeling to reduce the required model order at low pitch frequencies.

3. FREQUENCY WARPED ALL-POLE MODELING

A property of LP spectral approximation is that it is equally accurate at all frequencies [3]. However human auditory perception has a resolution that decreases with frequency. If the model order is low, it is possible that there is a preservation of spectral details at higher frequencies at the cost of accurate modeling of the spectral envelope at the perceptually more important lower and middle frequencies. A suitable warping of the frequency scale so as to transform the spectral envelope into one in which the lower frequency regions now occupy a larger portion of the frequency range while the higher frequency regions are correspondingly compressed has the potential to result in perceptually more accurate LP spectrum matching. The idea of frequency warping in context of linear prediction was applied by Makhoul [3], Strube in [9] for ASR and later by Koljonen et. al. [10] in speech coding. Later Harma et. al. applied it in context of audio signals at different sampling rates [11]. A warping of frequency scale may be achieved by applying a transformation [12] such as

$$\theta = f(\omega) = \arctan\left(\frac{(1-\alpha^2)\sin(\omega)}{(1+\alpha^2)\cos(\omega)+2\alpha}\right) \qquad (1)$$

where the uniformly spaced L harmonic frequencies $\{\omega_1, \omega_2, ..., \omega_L\}$ are mapped to a warped scale given by $\{\theta_1, \theta_2, ..., \theta_L\}$. The parameter α controls the severity of warping. Parameter $\alpha = 0.4$ approximates the auditory Bark scale at 8 kHz sampling frequency [13].

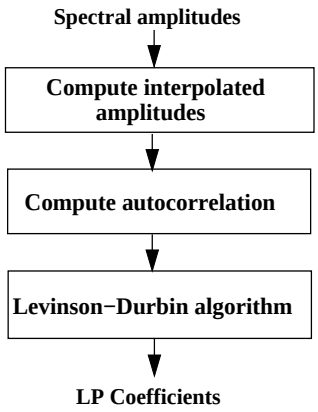

Figure 1: Frequency domain approach to all-pole modeling

3.1. The various perceptual scales

By varying the warping parameter in Eq 1 it is possible to approximate the different perceptual scales mentioned in the literature. Figure 2 illustrates a com-

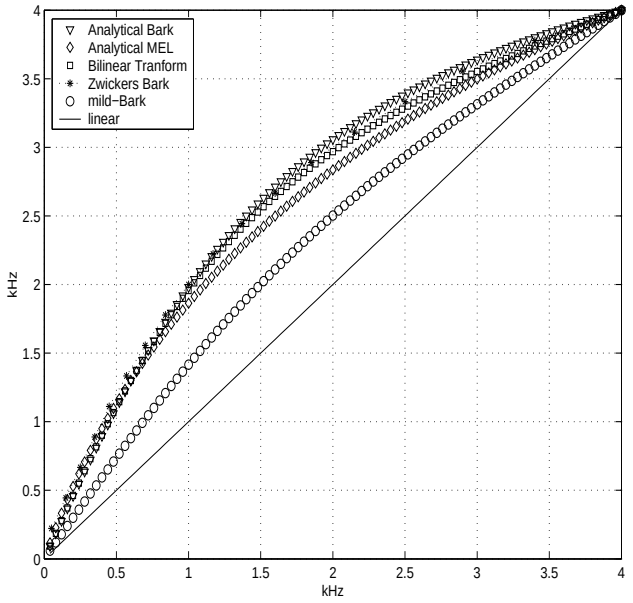

Figure 2: Comparion of various perceptual scales

parison of various perceptual scales. The MEL scale [14] may be approximated by the warping parameter of $\alpha = 0.3$. Figure 2 also shows a close match between the published Bark scale (Zwicker's Bark scale [13]), bilinear transform based Bark scale and analytical expression [14] that approximates the Bark scale. We also illustrate two more mappings that correspond to no warping (or conventional LP modeling) and a mild-Bark scale warping ($\alpha = 0.2$). For a fixed warping parameter ($\alpha = 0.4$ in case of Bark scale) a fixed band of frequencies on y-axis corresponds to a band of almost same width on x-axis in low frequency region. However at higher frequencies, the same width of frequencies on y-axis corresponds to a wider band of frequencies on x-axis. This matches with the definition of critical band scale, where a band of frequencies on a linear frequency scale maps to a constant distance on the basilar membrane (BM). As the frequency increases, wider and wider band of frequencies on linear scale maps to same width on the BM.

3.2. Implementation

The spectral amplitudes are obtained from frequency domain analysis of 20 ms windowed speech segments by an analysis-by-synthesis method based on the DFT [15]. The output of the analysis is a set of estimated amplitudes (or the reference amplitudes) $\{S(\omega_1), S(\omega_2), ..., S(\omega_L)\}$ at the uniformly spaced L harmonic frequencies $\{\omega_1, \omega_2, ..., \omega_L\}$. The harmonic frequencies are mapped to another set $\{\theta_1, \theta_2, ..., \theta_L\}$ of

Vowel	**Typical word**	**Pitch range**	IPA Symbol
/uh/	"but"	101 Hz - 131 Hz	ʌ
/a/	"guard"	97 Hz to 126 Hz	ɑ
/ow/	"law"	95 Hz - 130 hz	ɒ
/ae/	"cat"	87 Hz - 140 hz	æ
/oo/	"boot"	100 Hz - 123 Hz	u
/iy/	"sleep"	100 Hz - 138 Hz	i

Table 1: Description of vowel sounds used in the subjective listening experiment

warped frequencies through Eq 1. The spectral amplitudes are then log linearly interpolated to a fixed frequency spacing of 20 Hz to get the interpolated spectrum as follows:

$$Q(\theta_j) = 10^{\log|S(\theta_k)| + (\frac{\theta_j - \theta_k}{\theta_{k+1} - \theta_k}) \log|S(\theta_{k+1})| - \log|S(\theta_k)|} \quad (2)$$

$for \quad \theta_k < \theta_j < \theta_{k+1}$

where, $S(\theta_k)$'s are the set of spectral amplitudes obtained from spectrum analysis and the $Q(\theta_j)$'s are the interpolated amplitudes. We thus obtain 200 spectral samples in the 4kHz speech bandwidth. The autocorrelation is computed from the power spectrum by IDFT operation, and given by

$$R_i = \frac{1}{200} \sum_{j=0}^{199} |Q(\theta_j)|^2 \cos(i\theta_j) \quad (3)$$

Finally, the warped LP coefficients 's are computed by solving the following simultaneous equations using the Levinson-Durbin algorithm

$$\sum_{k=1}^{p} a_k R_{|i-k|} = -R_i \quad (4)$$

$for \quad 1 \leq k \leq p$

where p is the order of all-pole model. The all-pole model coefficients represent the spectral envelope. The spectral amplitudes are later recovered for speech synthesis by generating the envelope as in Eq 5 below and sampling it at warped frequency locations.

$$\hat{S}(\theta_i) = \frac{G}{1 + \sum_{k=1}^{p} a_k e^{-j\theta_i k}} \quad (5)$$

The $\{\hat{S}(\theta_i)\}$ are the modeled spectral amplitudes. The warping compresses the higher frequencies thus introducing greater inaccuracies in the modeled spectral amplitudes of the high frequency region.

4. SUBJECTIVE AND OBJECTIVE EVALUATION

In order to study the role of frequency warping, we consider a test set of sounds across vowel and voice qualities. Since LP modeling of the spectral magnitudes is most challenging for low pitched voices, we consider only male voices. Voice quality in context of speech analysis/synthesis refers to voice type and is largely determined by the glottal source excitation. Normal male voiced phonemes are classified as: modal, vocal fry (creaky) and breathy [16]. Each voice quality results due to acoustic characteristic of the underlying glottal excitation spectrum. Modal voice quality is a result of very sharp glottal closure, while breathy voice results from relatively long glottal closure phase and incomplete or poor contact of vocal folds [16]. Spectra of breathy voices show a strong first harmonic and steep spectral slope as compared to modal voice. Moreover due to incomplete glottal closure they may contain random noise.

In the context of the study presented in the paper, i.e. spectral envelope modeling, we have chosen the voice qualities that can be distinguished based on slope of the spectrum. We here on refer to the sound quality resulting from the flatter excitation spectrum as modal-*sharp* and the one due to steeper excitation spectrum as modal-*dark*. The vowel sounds, corresponding to one of the two voice qualities modal-*dark* and modal-*sharp*, were obtained as described later. The voice quality was distinguished based on subjective judgment and visual observation of relative strength of first harmonic with respect to the remaining harmonics. Figure 4a and Figure 6a illustrate modal-*dark* and modal-*sharp* spectra. We can observe that the modal-*dark* /iy/ (Figure 4a) has weaker higher formants F2, F3 and F4 than the modal-*sharp* /iy/ (Figure 6a). The subjective percept used for distinguishing the vowels with differing vowel qualities was *sharpness*. Modal-*sharp* voices, because of relatively strong higher harmonics are perceived to be significantly sharper (brighter) than their modal-*dark* counterparts.

Thus we had two sets of the same six vowel sounds, one corresponding to modal-*dark* voice quality and other corresponding to a modal-*sharp* voice quality. The vowel segments were manually extracted from individual words taken from the TIMIT database as well as sentences recorded in our lab. Utterances by 12 different adult male speakers were used to extract the vowel sounds. At times we found the extracted vowel sounds were too short, in such cases the period at start and end of the steady portion of the vowel was repeated to generate a vowel with longer duration. The ends were tapered to eliminate abrupt transitions. The duration of the sounds ranged between 400 ms and 700 ms. The set of representative vowel sounds used for our study are shown in Table 1. Table 2 and Table 3 list the actual vowel sounds used in subjective tests. Vowels illustrated in Table 2 are all modal-*dark* quality vowels while those in Table 3 are all modal-*sharp* quality vowels.

Vowel	Instance	Objective Ranks			Subjective Ranks			Speam. Corr.
		U	M	B	U	M	B	
/a/	AF1	3	2	1	3	2	1	1
	AF6	3	2	1	3	2	1	1
/ae/	AEF1	3	1	2	3	2	1	0.5
	AEF2	3	2	1	3	1	2	0.5
/uh/	UHF1	3	2	1	3	2	1	1
	UHF4	3	2	1	3	1	2	0.5
/ow/	OWF3	3	2	1	3	2	1	1
	OWF4	3	2	1	3	2	1	1
/iy/	IYF2	2	1	3	1	2	3	0.5
	IYF4	1	2	3	1	2	3	1
/oo/	OOF1	3	2	1	1	2	3	-1
	OOF4	3	2	1	1	2	3	-1

Table 2: Subjective and objective ranks for modal-*dark* quality vowel sounds

The sounds were analyzed to estimate the pitch and spectral amplitudes as described in 3.2. The spectral amplitudes thus obtained were modeled for each frame using 10th order frequency-warped LP modeling with a chosen warping factor (refer section 3.2). Synthesis was carried out by standard sinusoidal synthesis methods [15] using the spectral amplitudes obtained from the all-pole model approximation and compared with a reference sound synthesized using the originally estimated spectral amplitudes. There were 3 test sounds for each reference sound: LP modeled without frequency warping (denoted "U"), LP modeled with mild-Bark scale warping ("M") and LP modeled with Bark warping ("B").

4.1. Subjective tests

Subjective listening tests were conducted with normal hearing subjects, where subjects were asked to rank the relative perceived degradations of the test sounds U, M and B with respect to the corresponding reference sound for each of the vowel sounds in test set. Five normal hearing subjects participated in the test.

The test material was presented to the subject at normal listening levels through high quality head phones connected to a PC sound card in a quiet room. Subjects were allowed to listen to the reference and test sounds any number of times before making a decision. A particularly convenient method of listening was found to be the "reference-test-reference" sequence. Each lis-

Vowel	Instance	Objective Ranks			Subjective Ranks			Speam.
		U	M	B	U	M	B	Corr.
/a/	AM2	1	3	2	1	2	3	0.5
	AM7	1	3	2	1	2	3	0.5
/ae/	AEM8	1	2	3	1	2	3	1
	AEM9	2	1	3	1	2	3	0.5
/uh/	UHM5	1	2	3	1	2	3	1
	UHM7	2	1	3	1	2	3	0.5
/ow/	OWM3	1	2	3	1	2	3	1
	OWM6	1	2	3	1	2	3	1
/iy/	IYM2	2	1	3	1	2	3	0.5
	IYM5	2	1	3	1	2	3	0.5
/oo/	OOM2	1	3	2	1	2	3	0.5
	OOM4	1	3	2	1	2	3	0.5

Table 3: Subjective and objective ranks for modal-*sharp* quality vowel sounds

tener did the test using the same set of items in different ordering on three separate occasions. Although no instructions whatever on the type of degradation to listen for were given to the listeners, it was observed by them that the distortions due to modeling inaccuracies are characterized by changes in both, the intelligibility (clearness) and the "color" (brightness) of sound. Table 2 and Table 3 provide the subjective ranks for modal-*dark* and modal-*sharp* voice qualities respectively. Rank=1 implies the least perceived degradation and rank=3 the most degradation. To make the subjects familiar with the differences in sound introduced by warping which they had to notice, a training sound sequence was provided to each subject before start of the subjective test in which the test sound is the result of various degrees of warping.

4.2. Objective measurements

In order to obtain an insight into the dependence of perceived quality on the spectral modifications obtained by frequency warped all-pole modeling, an objective estimation of the perceived distortion was attempted using a psychoacoustically based distance measure known as "partial loudness". That is, the modeling error is treated as the signal whose audible significance is to be estimated in the presence of a background masker (the reference sound). This loudness of the error in presence of background masker, can serve as a psychoacoustically sound measure to quantify the degradation caused due to modeling errors.

Partial loudness (P.L.) was first applied to measure speech quality degradation due to quantization in [17]. Recently a computational model of partial loudness was proposed that accounts for a large body of subjective data from psychoacoustical experiments [18]. This model is based on the approximate stages of auditory processing representing the conversion of an input sound spectrum to the excitation pattern on the basilar membrane. In the case of a signal presented with a background masker, a partial loudness is derived for the signal based on the computed excitation patterns of the signal and the masker. The partial loudness model was shown to perform well in the prediction of audible discrimination of spectral envelope distortions in vowel sounds as measured in a psychoacoustical experiment [19]. In the present study, however, we are concerned with the ranking of degradations all of which are clearly audible (supra-threshold distortions). The reference sound is intended to take the role of the background noise and the modeled sound that of the signal plus background noise (reference sound plus the modelling error). The linear spectral distortion is treated as additive noise with power spectrum given by difference between reference and modeled power spectra.

The objective rankings of relative degradation were derived by computing the distortion measure for each of the reference-test (U/M/B) sound pairs. The ranks obtained by the objective measures are shown in Table 2 and Table 3. The performance of an objective distance measure in predicting subjective judgments may be evaluated by computing a measure of correlation between the objective rankings and subjective rankings. The Spearman's correlation coefficient [20] is a suitable measure since it makes minimal assumptions about the data. Applying the Spearman's correlation coefficient to the results of Table 2 and Table 3, we find that PL shows a positive correlation (equal to either 1.0 or 0.5) for all the test items except the vowels /oof1/ and /oof4/.

5. DISCUSSION

Table 2 illustrates effect of warped all-pole modeling on the vowels having modal-*dark* voice quality. We can observe that the rank 1 appears predominantly in the column marked 'B' indicating accuracy of all-pole modeling is improved with introduction of warping function. We can notice that the instances of vowel /iy/ and vowel /oo/ have actually shown the reverse trend. We can observe in Table 2 the vowels /uh/, /a/, /ae/, /ow/ improve with warping while the vowels /iy/ and /oo/ degrade after Bark scale warping. From Table 3, we see that all the vowels with modal-*sharp* sound quality degrade after Bark scale warping function is introduced in the all-pole modeling. Table 3 also indicates vowels /iy/ and /oo/ degrade after Bark scale warping. The vowel quality of /iy/ and /oo/ influences the perceived modeling error and irrespective of voice quality this vowel degrades after Bark scale warping is introduced. We see therefore that the audible sig-

nificance of the modeling error in case of Bark scale warped all-pole modeling of vowel sounds are largely influenced by the spectral content.

Figures 3 to 6 illustrate how the mismatch between the estimated (original) $S(.)$ and modeled $\hat{S}(.)$ spectral amplitudes for each of two vowel sounds translate into a partial loudness distribution on the auditory ERB-rate scale (an auditory scale related to the frequency in Hz through an approximately logarithmic relation). The P.L. objective distance is the integral of the partial loudness distribution. Figure 3 and Figure 5 illustrate the comparison of original and modeled spectral envelopes and the specific loudness of modeling error for vowel /a/ for two types of voice quality for unwarped and Bark warped case. The strength of higher harmonics is relatively high for modal-*sharp* /a/, and any modeling error introduced by the Bark scale warping becomes audibly significant. This suggests that the speaker variability plays important role in influencing the perceived modeling error. Figure 4 and Figure 6 illustrate comparison of original and modeled spectral envelopes and the specific loudness of modeling error for vowel /iy/ corresponding to the two voice qualities for unwarped and Bark warped conditions. In both the cases the Bark scale warping makes the modeling error audibly more significant.

While frequency-warped LP modeling with high warping parameters (such as the Bark scale) universally improves the spectral match in the low frequency region, the spectral envelope errors in the high frequency region vary in their contribution to the perceived distortion. How significant this high frequency spectral distortion is depends not only on its exact frequency location but also the extent of frequency masking provided by neighboring spectral components. As compared to vowels /a/, /ae/, /ow/ and /uh/, vowel /iy/ is characterized by relatively widely spaced first and second formant. The first formant is located around 400 hz. This results in a trough in the spectral envelope in the 500 Hz to 1500 Hz. The harmonics lying in between the two formants, the first and second, have relatively low strength and produce insufficient excitation to mask the modelling error, caused due to frequency warping. Thus in case of /iy/ there is significantly less upward spread of masking from low frequency components compared with that in /uh/. We noted that for some instances modal-*dark* /oo/ vowel improved with frequency warping, this was observed for those samples of /oo/ in which, the second, third and fourth formants were not clearly distinguishable. While in most instances where the higher three formants could be distinguished, degradation could be noticed after frequency scale warped all-pole modeling. We have found that both "high" vowels (/iy/ and /oo/)typically degrade under Bark-scale warped spectral modeling. Our observations on isolated vowels have been borne out also in informal listening to sentences containing the predominance of one or another vowel.

Out of the total of 24 instances that we considered in Table 2 and Table 3, the mild-Bark warped test sound has been ranked as best or second best almost all the time. The modeling error due to mild-Bark scale warping function is relatively robust to vowel quality or speaker variability, as compared to the modeling error due to unwarped or Bark warped conditions. This suggests that mild-Bark scale warping function (corresponding to warping parameter $\alpha = 0.2$) is a good candidate as "universal" warping function.

6. CONCLUSION

We have considered frequency warped all-pole representation of discrete spectral amplitudes of vowel sounds generated by sinusoidal models in the context of low bit rate speech coding. LP modeling of the spectral envelope has been the most popular method of quantizing the spectral envelope. In low bit rate speech coding applications, where a low LP model order is desirable, it is necessary to know the factors that influence the modeling error. It is found that low pitched sounds tend to degrade more than high pitched sounds at low LP orders. It has been widely noted in the literature that Bark-scale frequency warped all-pole modeling improves the accuracy of modeled spectral envelope. However, our study suggests, that the shape of spectral envelope plays important role in influencing the modeling error at low model orders. The spectral shape of the vowel sounds is determined by vocal tract transfer function (which determines vowel quality) and the glottal excitation (which determines the overall spectral slope). Our subjective tests on a set of vowels across different speakers suggest that the vowels having relatively flat spectra (resulting from a flatter excitation spectrum) degrade when modeled with Bark scale frequency warped all-pole modeling, for such sounds Bark scale frequency warping should be avoided. Also vowels, such as /iy/ and /oo/ have an inherent spectral shape which is responsible for degrading the quality of such Bark warped all-pole modeled vowel sounds. Based on the results of the computation of an auditory distance measure we find that the presence of relatively low strength harmonics in the lower frequency region generates a masking threshold that is insufficient to conceal the audibility of high frequency modeling errors introduced by the the Bark scale frequency warping. It is therefore necessary to use the

Bark scale frequency warping judiciously.

If a single frequency warping function is to be used in all-pole modeling, our study suggests that mild-Bark scale warping function, is more robust choice and can be used universally. Study is underway to investigate suitable metrics to predict whether frequency warping would be beneficial or not.

Acknowledgment: *Detailed comments from an anonymous reviewer were greatly appreciated.*

REFERENCES

[1] MacAulay R.J. and Quatieri T.F., "Sinusoidal coding," in Speech Coding and Synthesis, Elsevier, Amsterdam, 1995.

[2] Das A. and Gersho A., "Variable dimension spectral coding of speech at 2400 bps and below with phonetic classification," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 492 –495, Apr 1995.

[3] Makhoul J., "Spectral linear prediction: Properties and applications," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, no. 3, pp. 283–296, Jun 1976.

[4] Hermansky H., Hanson B.A., Wakita H., and Fujisaki H., "Linear predictive modeling of speech in modified spectral domain," in *Digital Processing of Signals in Communications*, Apr 1985, pp. 55–63.

[5] El-Jaroudi and Makhoul J., "Discrete all-pole modeling," *IEEE Transactions on Signal Processing*, vol. 39, no. 2, pp. 411–423, Feb 1991.

[6] B. Wei and Gibson J.D., "Comparison of distance measures in discrete spectral modeling," in *Proc. of IEEE Digital Signal Processing Workshop*, Oct 2000.

[7] Patwardhan P. and Rao P., "Controlling perceived degradation in spectrum envelope modeling via predistortion," in *7th International Conf. on Spoken Language Processing ICSLP 2002*, Denver, USA, Sep 2002, pp. 1837–1840.

[8] Champion T.G., McAulay R.J., and Quatieri J.F, "High-order all pole modeling of the spectral envelope," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Apr 1994, pp. 529–532.

[9] Strube H. W., "Linear prediction on a warped frequency scale," *J. Acoust. Soc. of Am*, vol. 68, no. 4, pp. 1071–1076, Oct 1980.

[10] Koljonen J. and Kajalainen M., "Use of computational psychoacoustical models in speech processing: Coding and objective performance evaluation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1984.

[11] Harma Aki and Laine U.K., "A comparison of warped and conventional linear predictive coding," *IEEE Trans. Speech Audio Processing*, vol. 9, no. 5, pp. 579–588, Jul 2001.

[12] Oppenheim A.V., Johnson D.H., and Steiglitz K., "Computation of spectra with unequal resolution using fast Fourier transform," *Proc. of IEEE*, vol. 59, no. 2, pp. 299–301, Feb 1971.

[13] Smith J. and Abel J, "Bark and erb bilinear transform," *IEEE Trans. Speech and Audio Processing*, vol. 7, no. 6, pp. 697–708, Nov 1999.

[14] Douglas O'Shaughnessy, *Speech Communications*, University Press, 2001.

[15] Griffin D.W.and Lim J.S, "Multiband excitation vocoder," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, no. 8, pp. 1223–1235, Aug 1988.

[16] Childers D.G. and Lee C.K., "Vocal quality factors: Analysis, synthesis and perception," *J. Acoust. Soc. Am.*, vol. 90, no. 5, pp. 2394–2411, Nov 1991.

[17] Schroeder M.R., Atal B.S., and Hall J.L, "Objective measure of certain speech signal degradations based on masking properties of human auditory perception," in Frontiers of Speech Communication Research, Academic NewYork, 1979.

[18] Moore B.C.J., Glasberg B.R., and Baer T., "Model for prediction of thresholds, loudness and partial loudness," *J. Audio Eng. Soc.*, vol. 45, no. 4, pp. 224–240, 1997.

[19] Rao P., van Dinther R., Veldhius R., and Kohlrausch A, "A measure for predicting audibility discrimination thresholds for spectral envelope distortions in vowel sounds," *J. Acoust. Soc. Am*, vol. 109, no. 4, pp. 2085 –2097, Apr 2001.

[20] Jane Miller, "Correlation," in Statistics for Advanced Level , Cambridge University Press, 1989.

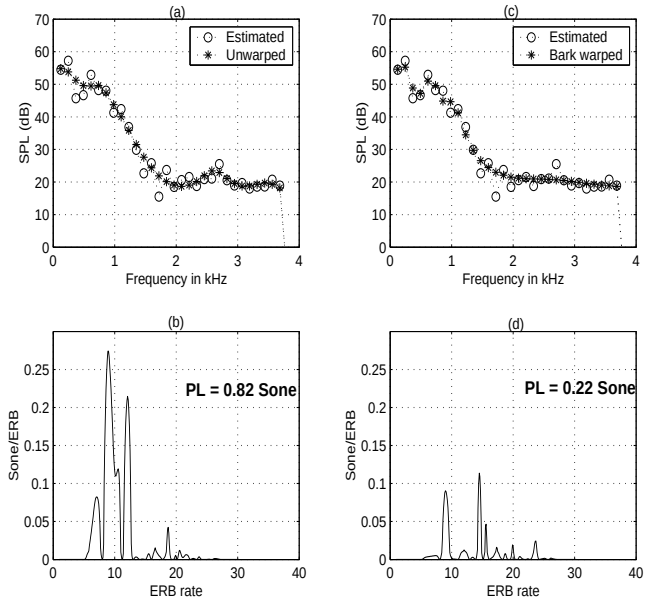

Figure 3: Effect of Bark-warped all-pole modeling on vowel /a/ (pitch = 123 Hz, LP-model order 10, typical word: "guard") which has modal-*dark* voice quality. (a),(c) Unwarped and Bark-warped LP spectral envelopes respectively. (b)(d) Specific loudness plots for unwarped and Bark-warped cases respectively

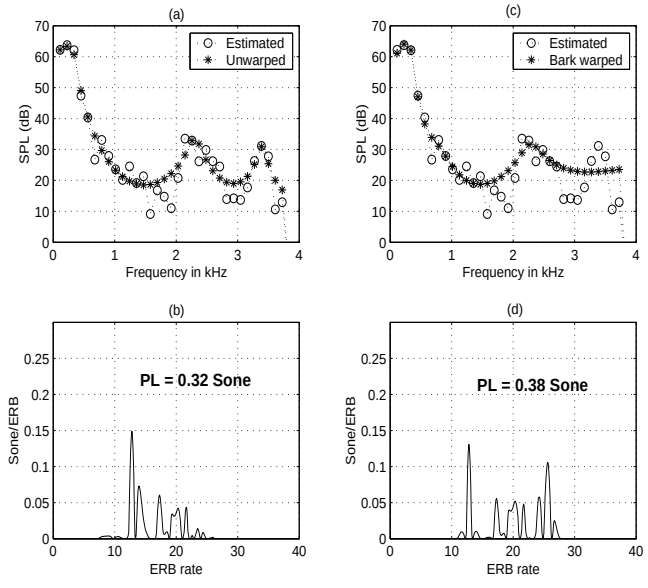

Figure 4: Effect of Bark-warped all-pole modeling on vowel /iy/ (pitch = 115 Hz, LP model order 10, typical word: "feet") having modal-*dark* voice quality. (a),(c) Unwarped and Bark-warped LP modeled spectral envelopes respectively. (b),(d) Specific loudness plots for unwarped and Bark-warped cases respectively

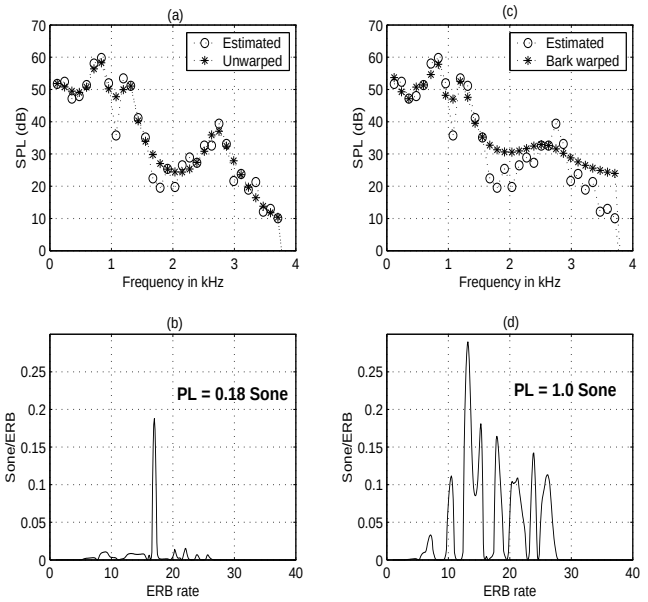

Figure 5: Effect of Bark-warped all-pole modeling on vowel /a/ (pitch = 119 Hz, LP-model order 10, typical word: "guard") which has modal-*sharp* voice quality(a),(c) Unwarped and Bark-warped LP modeled spectral envelopes respectively. (b),(d) Specific loudness plots for unwarped and Bark-warped cases respectively.

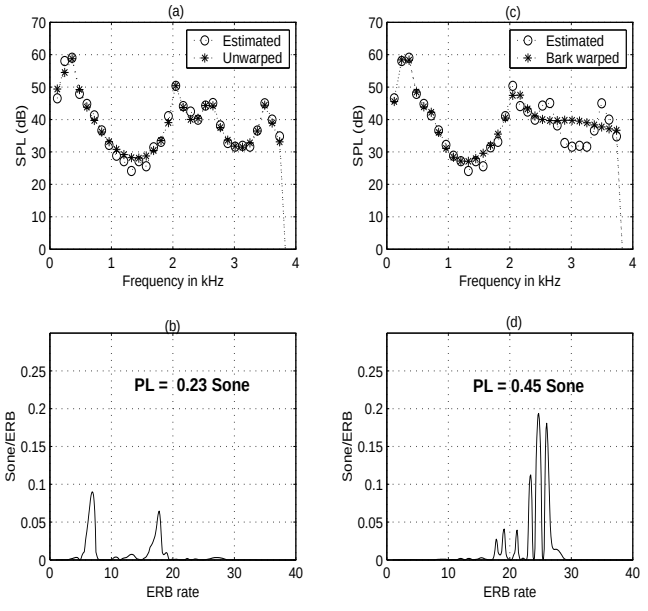

Figure 6: Effect of Bark-warped all-pole modeling on vowel /iy/ (pitch = 100 Hz, LP model order 10, typical word: "feet") having modal-*sharp* voice quality. (a),(c) Unwarped and Bark-warped LP modeled spectral envelopes respectively. (b),(d) Specific loudness plots for unwarped and Bark-warped cases