SYSTEMATIC COMPARISON OF SPOKEN AND SUNG VOWELS USING PERCEPTUAL LINEAR-PREDICTION ANALYSIS

Thomas J. Millhouse
Sydney Conservatorium,
University of Sydney,
Sydney, Australia
Thomas.Millhouse@defence.gov.au

Frantz Clermont
School of Mathematics, Science and Computer Science
American University of Paris,
Paris, France
frantz.clermont@aup.fr

Abstract

This paper introduces the use of perceptual linear prediction (PLP) analysis to observe systematic differences in spoken and sung vowels in the acoustic-auditory domain. PLP is a recent evolution of speech acoustic analysis that models more closely the psychophysics of hearing to derive an estimate of the auditory spectrum, and accounts for the effective second formant and 3.5-Bark spectral-peak integration theories of vowel perception. In the acoustic domain, the sung vowel introduces a clustering and lowering of the higher formant structure, a phenomenon called the singer's formant. In the auditory domain, the singer's formant has an observable effect on the frequency and bandwidth of the effective second formant, which in turns raises questions on some of the presently accepted theories on the perceptual nature of vowels and the contributions of auditory spectral regions to vowel classification.

1. Introduction

The study of the singing voice is influenced strongly by the academic literature of acoustic speech science and singing pedagogy. One body of knowledge draws its conclusions on the singing voice from pure acoustical analysis, whilst the other lends towards a perceptually oriented model for examination. The evolution of speech acoustic analysis into an auditory specific domain via perceptual linear prediction (PLP: Hermansky 1990), presents itself as the ideal analytical instrument from which to compare the observations of both bodies of knowledge together.

Of particular interest to the speech scientist and the vocal pedagogue are the differences in formant frequencies between spoken and sung vowels. Formant frequencies have had a long tradition in speech analysis, due to the important information that they represent on the shape of the vocal tract, and their perceptual significance relating to phonetic differentiation and speaker-dependant qualities. Acoustic analysis detailing the formant structure of sung and spoken vowels has revealed that in singing the unrounded front vowels are shifted towards the back vowels (Clermont 2002) and the upper formant structure is lowered and clustered in the singer's formant region (Sundberg 1974). In bass and baritone voices, acoustical formant analysis is relatively simple via linear prediction techniques; however, the wide spacing of harmonics in high-pitched singing has made the evaluation of acoustic formant frequencies problematic and unreliable in the higher pitched voices. It remains apparent that a gap remains in the knowledge of the formant structure of the singing voice in the acoustic domain, due to these analytical difficulties.

In comparison to their acoustic domain counterparts, the discussion of formant frequencies in the perceptual domain is a relatively newer concept. Although the learned scientist would state that Sir Isaac Newton first documented a concept similar to the perceptual formant in 1665 (Ladefoged 2001, Pg 173), the notion that the human ear might be performing spectral reduction was more formally recognised by Chiba & Kajiyama (1941) and again by Delattre *et al* (1954). Both these studies indicated that only two spectral peaks are required for simulating the phonetic qualities of front vowels whilst only a single peak is required for back vowels. Fant & Risberg (1962) took these studies further by incorporating synthetic simulations of Swedish vowels to determine a perceptual formant frequency determined by subjective listener experiments. They proposed that by keeping the first perceptual formant frequency at its acoustic first formant equivalent, the *effective second formant* F_2' or the perceptual second formant required for phonetic differentiation does not directly correspond to any particular formant frequency. F_2' was found to be typically close to the second formant frequency in simulating back vowels and close to the third and fourth frequency in front vowels.

Perceptually oriented acoustic studies of the singing voice have had more success in detailing differences in spoken and sung vowels, then their pure acoustical counterparts. Bloomploft & Plomp (1985) studied the differences in spoken and sung vowels using a 1/3 octave filter bank and a statistical process of known as Principal Component Analysis (PCA). This model delivered a two-parameter description of perceptual phonetic discrimination in the sung voice, based on the two highest principal components, which in turn were demonstrated to be representative of the acoustic formant frequencies.

The analytical capabilities of perceptual linear predictive analysis by Hermansky (1990) have made the approximation of perceptual formant frequencies from acoustic signals a closer reality. Perceptual Linear Prediction analysis based on a five coefficient model (PLP5) (Hermansky 1990) derives two perceptual formant frequencies that closely match the results of subjective listener experiments (Fant & Risberg 1962). The PLP5 technique therefore holds an advantage over the statistical PCA model (Bloomploft & Plomp 1985) in that perceptual formants can be derived for a single vowel at a time. In addition, the advent of the PLP5 approach to speech analysis has shed light on a number of not yet fully accepted concepts of speech production and perception. Two particular psycho-acoustical topics yet to be fully quantified, are; the effective second formant F_2' (Fant and Risberg, 1962), and the 3.5 Bark spectral-peak integration theory (Chistovich *et al* 1978). These theories of vowel perception hold significant interest for the perceptual analysis of the sung vowel. Initially the clustering of the higher formant structure

Proceedings of the 10th Australian International Conference on Speech Science & Technology
Macquarie University, Sydney, December 8 to 10, 2004. Copyright, Australian Speech Science & Technology Association Inc.

Accepted after full paper review

would imply from the theories of Chistovich *et al* (1978) that the singer's formant, as defined by Sundberg (1974), would be perceived as a single dominate perceptual formant due it its increased spectral intensity. This dominate perceptual formant would then in turn effect the position of F_2', according to Fant and Risberg (1962), however it is unknown whether or not this altered position of F_2' would therefore modify the perceptual discrimination of the sung vowel.

The principal argument for this paper is that the singer's formant, a supposedly speaker dependant parameter, effects the perception of the sung vowel in the auditory-acoustic domain. The paper will initially introduce the PLP5 method of acoustic analysis for the derivation of the sung auditory domain and the comparison of spoken and sung vowels there within. Then through the use of a machine based classification method, this paper will discuss the role that the singer's formant has for the above-mentioned psycho-acoustical theories of vowel perception, focusing primarily on the role of the singer's formant in the perception of the phonetic quality of the sung vowel.

2. Materials and methods

2.1 Materials

To observe the acoustic-auditory characteristics of spoken and sung vowels, we used a complete set of Australian English vowels spoken and sung by the same male subject. He is a part-time chorus artist with a professional Australian opera company, and a native speaker of Australian English. He has had several years of training in western classical opera singing, has won regional Australian vocal championships, and regularly performs as a professional chorus artist and in minor operatic roles. Initially he spoke five randomized tokens of 11 monosyllables in /hVd/ context, at his habitual speaking rate (F_0=80Hz on average). Then queued with a 110Hz tone our subject sang the same tokens at approximately the same pitch as the queued tone. The analogue signals were then sampled at 11,025 Hz and quantized to eight bits.

2.2 Linear predictive formant extraction

To obtain the acoustic formant frequencies an analysis-by-synthesis technique in accordance with Clermont (1992) was employed. Formants were extracted from the poles of the LP analysis through (Hanning) windowed frames of 30 msec duration, by steps of 10ms. For 10% of the spoken data the LP-order was increased to 16 from a default value of 14 and 20 for 20% of the sung data, in order to enhance the upper formant regions. For each steady-state section, the analysis-by-synthesis technique yielded four sets of formants for each sung and spoken vowel.

2.3 Perceptual linear predictive formant extraction

The analytical process of perceptual linear prediction (PLP) is discussed extensively in Hermansky (1990). It determines an auditory spectrum based on natural speech perception by implementing observed psychoacoustic transforms relating to frequency resolution and the dependence of loudness on frequency and intensity. In this paper, the auditory spectrum was defined on the bark scale by effectively bandpass filtering the spoken or sung waveform into 19 critical bands over the frequency range of 0 to 11kHz. These bands were then further processed by Hermansky's (1990) equal-loudness preemphasis and cubic-root amplitude compression prior to the derivation of autoregressive coefficients.

Of particular importance in the PLP procedure is the choice of the number of autoregressive coefficients. Hermansky (1990) discussed at great lengths the capability of the higher order PLP procedure to carry both the linguistic message and speaker-dependant information, whilst lower order PLP analysis suppressed much of the speaker-dependant characteristics of a speaker.

> *The advantage of the PLP technique over the conventional LP is that it allows for the effective suppression of the speaker-dependent information by choosing the particular model order.*
>
> *- Hynek Hermansky (1990)*

As this paper sought to understand information relating to both the speaker-dependant and speaker-independent characteristics of the singing voice, a PLP analysis of 5 and 14 coefficients were employed side by side (Hence cited as PLP5 and PLP14). Specific perceptual formants were obtained by hand from the poles of the autoregressive modelling of the auditory spectrum, and were verified by comparison with the auditory spectrogram.

2.4 Perceptual formant determination from acoustic formants

In addition to the perceptual formants obtained by the PLP5 technique, as a confidence check perceptual formants for this subject were also determined as a function of the first four formant frequencies, in accordance with the formulae of Bladon and Fant (1978; hence cited as BF78). The formulae of BF78 produces a continuous shift of F_2' between the extreme frequencies of F_2 and $(F_3F_4)^{1/4}$.

$$F_2' = \frac{F_2 + c^2 (F_3 F_4)^{1/2}}{1 + c^2} \qquad c = K(f)\frac{A_{34}}{A_2} \qquad (1)$$

Where A_{34} is the vocal tract transfer function between F_3 and F_4 at the frequency $F_{34} = (F_3F_4)^{1/2}$ and A_2 is the transfer function at the second formant peak F_2. The factor K(*f*) in the weighting function is intended to compensate for additional pre-emphasis originating from the source, radiation and higher pole correction, and in addition a correction for differences in equal loudness level. For this experiment the K(*f*) was maintained at 12 in accordance with BF78.

3. Results

3.1 Vowel Spectra

To compare the spectral differences between the traditional acoustic LP magnitude scale and the auditory PLP5 and PLP14 scale, the average log magnitude spectra of all three analysis techniques was calculated for each vowel and displayed in the tabulated form of Figure 1. The spectrums are calculated over 5.5kHz and 19 Bark respectively whilst the amplitude were independent measures of dB, chosen to give the best spectral view of each method. Graphs of the average log magnitude spectra of each vowel for LP, PLP14 and PLP5 are recorded as Figure 1.

 Copyright, Australian Speech Science & Technology Association Inc.

Accepted after full paper review

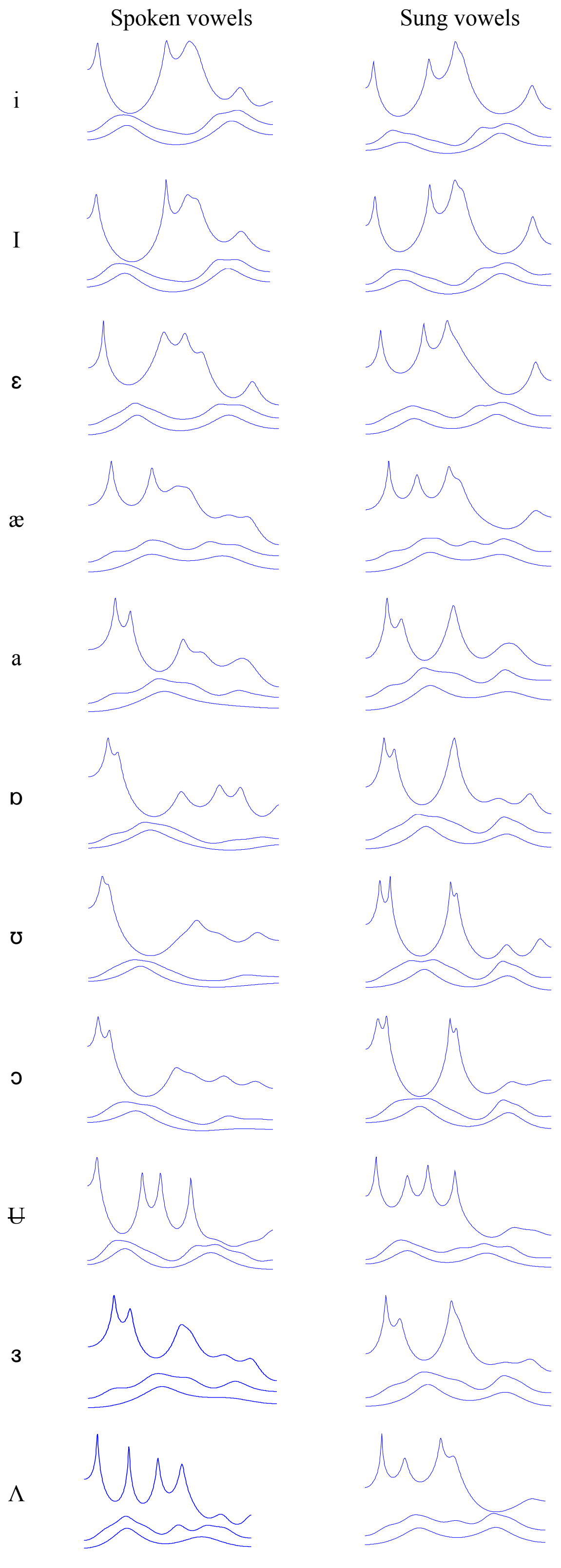

Figure 1: Comparative plots of the 11 spoken and sung vowels, comparing a standard 14th order LP on a Hertz scale, a 14th order PLP on a bark scale, and a 5th order PLP on a bark scale.

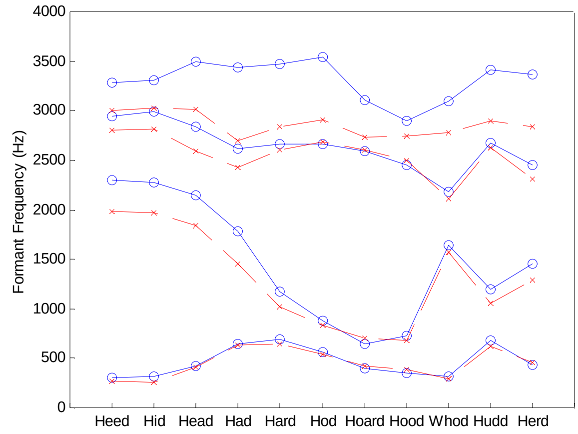

Figure 2: Acoustic Formant patterns (5 token averages) of Australian English vowel nuclei spoken and sung in /hVd/ context. Circles represent the spoken data whilst crosses represent the sung data.

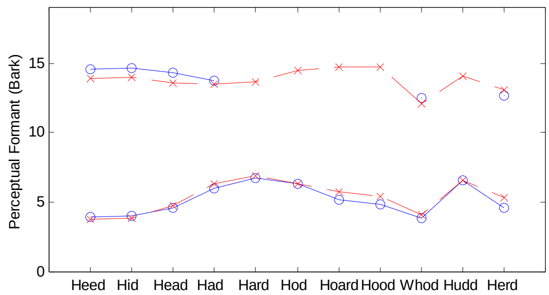

Figure 3: Perceptual formant patterns derived from the PLP5 analysis process (5 token averages) of Australian English vowel nuclei spoken and sung in /hVd/ context. Circles represent the spoken data whilst crosses represent the sung data.

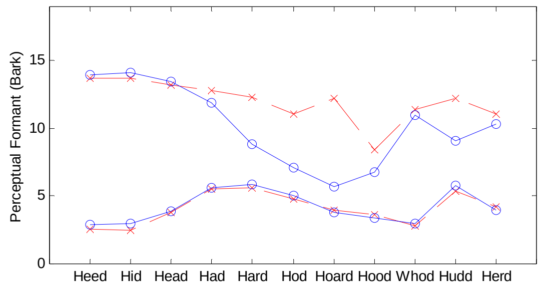

Figure 4: Perceptual formant patterns derived from the BF78 analysis process (5 token averages) of Australian English vowel nuclei spoken and sung in /hVd/ context. Circles represent the spoken data whilst crosses represent the sung data.

3.2 Vowel sequence charts

Vowel sequence charts for both the acoustic and perceptual formant domains are detailed in Figures 2-4. Figure 2 is the classic sequence chart demonstrating changes in acoustic formant position between the spoken and sung vowels for each of the eleven formants. Our acoustic formant tracking shows the decrease in F_2 for sung front vowels and the lowering and clustering of F_3 and F_4 around 2.8 kHz. These results are in general agreement with Sundberg (1970).

Figures 3 and 4 indicate the movement of the perceptual formants in the acoustic-auditory domain, using the PLP5 and

Macquarie University, Sydney, December 8 to 10, 2004. Copyright, Australian Speech Science & Technology Association Inc.

Accepted after full paper review

BF78 analysis techniques. Unlike their acoustic formant counterparts, only two spectral peaks are recorded and detailed for both the spoken and sung vowels. In PLP5 analysis, where the F_2' was undefined due to its merging with F_1', it was omitted from plotting. The two sequence charts reveal similar trends in perceptual formant movements for the formant detection methods in the acoustic-auditory domain. The perceptual tracking of both techniques indicates that in the auditory domain there is an increasing of F_2' in the back vowels towards that of the front vowels. This is the opposite result to observation of the lowering of F_2 in front vowels in the acoustic domain as discussed by Clermont (2002).

3.3 Tabulated effective second format

The final set of data for this experiment compares the F_2' predictions of BF78 with that of the PLP5 method. For our spoken data, we see that the results for the Australian English data are roughly in accordance with that of Hermansky's (1990) Swedish cardinal vowels. The positions of the perceptual peaks correspond closely with the F_1 and F_2' of the respective vowels, as determined by BF78. The PLP5 method clearly indicates the integration of F_2' into the F_1' in the case of back vowels, to form a single integrated peak in accordance with Chistovich (1978), whilst the BF78 model indicates that the predicted F_2' is closer than 3.5 Bark from F_1 once again in accordance with Chistovich (1978).

The sung data has some considerable differences though for the prediction of F_2'. Initially both the BF78 model and the PLP5 model present results that should resolve an F_2' for all sung back vowels (as opposed to merging with F_1' as in the spoken back vowels), however the methods do not agree on a common position for this F_2'. The BF78 model delivers F_2' results that are consistently lower than the PLP5, but both methods imply that a second perceptual formant is present for sung back vowels. (It should be noted though that Clermont (2002) indicated that the sung F4 was unreliable with his tracking methodology, and so the BF78 method may well be delivering unreliable results for the sung vowels and is intended only as a guide for observing formant movement.) The observation of a second perceptual formant being required for phonetic discrimination in sung back vowels is in disagreement with the multitude of academic research that indicates that the back vowels should by auditory discriminated from the front vowels by the presence of only a single perceptual peak.

A clear observable difference between each of the spoken and sung vowels from both the PLP5 and BF78 results, is the presence of a constant F_2' at approximately 14 Bark. This peak is most likely the result of the underlying F_3-F_4 clustering in the singer's formant region as predicted in the introductory section. The presence of this peak however presents itself as perceptually ambiguous, according to the presently understood theories of human psychoacoustics. For the spoken back vowel, only a single perceptual peak is required for phonetic discrimination, however in the sung back vowel a second peak is present potentially confusing the back vowel with a similar front vowel? An obvious research question arises from the observable differences in the effective second formant (F_2') between spoken and sung vowels. Does the sung F_2' influence the perceptual discrimination of the sung vowel?

4. Discussion

The presently accepted psychoacoustic theory on the effective second formant is that in the auditory domain, only two formants are required to simulate front vowels and only a single perceptual formant is needed to simulate back vowels (Fant & Risberg 1962). Our results have indicated that the F_2' appears to be relatively constant in singing calling into question its validity as a phonetically discriminating parameter on which the human auditory system differentiates the phonetic difference between vowels.

Table 1: Perceptually estimated (Bladon & Fant 1978) and PLP estimated frequencies of perceptual formants of the 11 Australian English spoken vowels.

	BF78			PLP5		Error	
	F_1	f_2'	f_2'-F_1	F_1'	F_2'	F_1'-F_1	F_2'-f_2'
Vowel	(Bark)			(Bark)		(Bark)	
i	2.9	14.0	11.1	4.0	14.6	1.1	0.5
I	3.0	14.1	11.1	4.0	14.6	1.0	0.5
ɛ	3.9	13.5	9.6	4.6	14.3	0.7	0.8
æ	5.6	11.9	6.3	6.0	13.7	0.5	1.9
a	5.9	8.8	2.9	6.8	NaN	0.9	NaN
ɒ	5.0	7.1	2.1	6.3	NaN	1.3	NaN
ʊ	3.8	5.7	1.9	5.2	NaN	1.4	NaN
ɔ	3.4	6.8	3.4	4.9	NaN	1.5	NaN
ʉ	3.0	11.0	8.0	3.9	12.5	0.9	1.5
ɜ	5.8	9.1	3.2	6.6	NaN	0.8	NaN
Λ	4.0	10.3	6.3	4.6	12.7	0.6	2.4

Table 2: Perceptually estimated (Bladon & Fant 1978) and PLP estimated frequencies of perceptual formants of the 11 Australian English sung vowels.

	BF78			PLP5		Error	
	F1	f2`	f2`-f1	F_1'	F_2'	F1`-F1	F2'-f2`
Vowel	(Bark)			(Bark)		(Bark)	
i	2.6	13.7	11.1	3.8	13.9	1.2	0.2
I	2.5	13.7	11.2	3.9	14.0	1.3	0.3
ɛ	3.8	13.2	9.4	4.8	13.6	0.9	0.3
æ	5.5	12.8	7.3	6.3	13.5	0.8	0.6
a	5.6	12.3	6.7	6.9	13.7	1.3	1.3
ɒ	4.8	11.1	6.3	6.3	14.5	1.5	3.3
ʊ	4.0	12.2	8.2	5.8	14.7	1.8	2.5
ɔ	3.6	8.4	4.8	5.4	14.7	1.8	6.4
ʉ	2.8	11.4	8.6	4.1	12.1	1.3	0.7
ɜ	5.4	12.2	6.7	6.6	14.1	1.1	1.9
Λ	4.2	11.1	6.9	5.3	13.1	1.1	2.0

4.1 Auditory spectral regions of primary variance

Although the effective second formant for sung vowels has been demonstrated to be reasonably stable it still may have an effect on the perception of the vowel by varying its bandwidth and spectral shape. Fujimura (1967) questioned the perceptual role of F_2' by indicating the existence of pairs of phonetically distinct vowels with identical F_1 and F_2', however such vowels differed in the spread of the higher formant structure. Bladon (1983) expanded the F_2' theory, and discussed the ambiguities of Fujimura (1967) to indicate that the bandwidth of F_2' was also important in phonetic discrimination as it held information on the clustering of the higher underlying format structure.

Proceedings of the 10th Australian International Conference on Speech Science & Technology
Macquarie University, Sydney, December 8 to 10, 2004. Copyright, Australian Speech Science & Technology Association Inc.

Accepted after full paper review

Bandwidth extraction from autoregressive poles of a regular acoustic spectrum is problematic and prone to error as discussed by Vallabha and Tuller (2002). So in order to determine if the spectral shape of the F_2' might play a role in the perceptual nature we look towards a measure of inter-vowel spectral variance in the auditory domain. Figure 5 shows the amount of inter-vowel variance calculated over six selected spectral regions of three bark intervals. Variance was calculated in each specific band region as the sum of the average individual vowel cepstra distances, from the average of all vowels cepstra, divided by the total number of vowels. Distance measurements were based on the Clermont and Mokhtari (1994) frequency specific distance measure transformed into the auditory domain. One can observe that vowel variance is greatest across the entire $F_1' – F_2'$ range, but within that range, the variance is greatest in the 12-15 bark range which corresponds to the F_2' range of influence. Interestingly the sung vowels have a greater range of variance in this region even though their formant centre frequencies appear to remain relatively stable. As the variance is greatest in the F_2' range but the centre frequencies remain stable the source of variance must reside within the bandwidth or shape of F_2' rather than its central frequency.

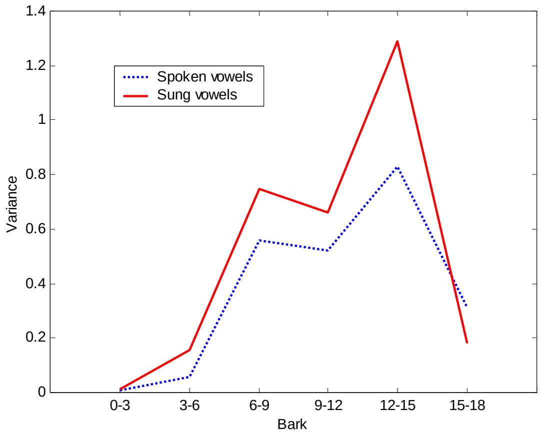

Figure 5: Spoken and sung vowel variance across selected auditory bandwidths.

4.2 Auditory spectral regions of primary phonetic influence.

To investigation the perceptual role of F_2' further we can apply a machine classification task to a limited auditory spectral band. Mokhtari and Clermont (1996; hence cited as MC96) introduced a method of determining the spectral regions of primary phonetic influence by using a band specific cepstrum distance measure. By transforming this method into the auditory domain, we can determine exactly what regions of the auditory spectrum are responsible for the majority of the phonetic classification.

The MC96 classifier chosen for this experiment is based on the nearest-neighbour method of pattern comparison. Four vowel tokens of the average auditory cepstra are used to determine the average auditory cepstra for the first four vowels. Then the frequency specific cepstra distance measure between the average cepstra and the fifth vowel token is used to find the closest match between the fifth token and the average cepstra of all the vowels (based on the first four tokens). The least distance measure will determine the vowel classification. By varying the auditory bandwidth (from 0 – 19 Bark), from which to carry out the distance measure, a measure of the vowel classification accuracy versus auditory spectrum bandwidth can be achieved.

Figure 6 demonstrates the vowel classification accuracy in the auditory domain for our band specific classifier. The results are for the all the spoken vowels, and then the front and back vowels are split into separate vowel groups for comparison. The spoken back vowels rise to a classification accuracy of 90% dependant only on having a band range of 0-6 Bark. This result agrees with Fant & Risberg (1962) in that spoken back vowels only require a single low spectral peak for perceptual identification. The front vowels have a much more even spread of vowel classification. To reach a similar level of accuracy as the back vowels they require a much larger auditory bandwidth, up to 17 bark in this case, for a 90% identification. This is indicative of the fact that they require the inclusion of both auditory formants for identification purposes, which once again in agreement with Fant and Risberg (1962).

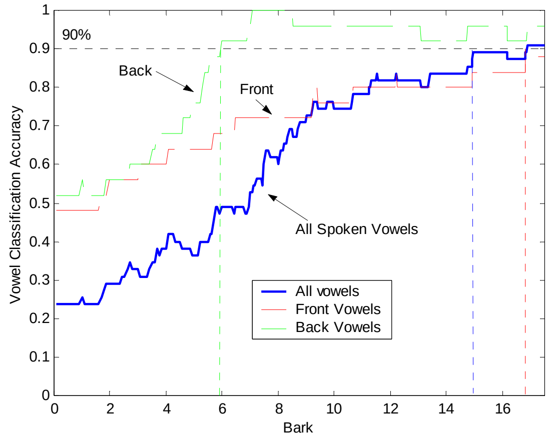

Figure 6: Vowel classification accuracy of selected auditory bandwidths for spoken vowels and spoken vowels separated into front and back vowels for classification.

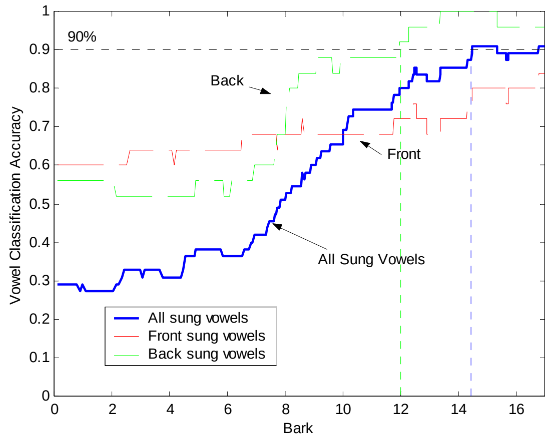

Figure 7: Vowel classification accuracy of selected auditory bandwidths for sung vowels and sung vowels separated into front and back vowels for classification.

Figure 7 looks at the same vowel classification accuracy for sung vowels. The results are once again separated into back and front vowels for comparison. The sung back vowel rises to a classification accuracy of 90% within a bandwidth range of 0-12 Bark, which is indicative of the fact that a greater spectral range encompassing both spectral peaks is required by the human ear to discriminate sung back vowels. In addition, the front vowels do not reach an accuracy of 90%

Macquarie University, Sydney, December 8 to 10, 2004. Copyright, Australian Speech Science & Technology Association Inc.

Accepted after full paper review

but rather increase gradually over the increasing auditory bandwidth at a similar rate to their spoken counterparts.

If we now turn our attention to a comparative observation of the vowel classification for both spoken and sung back vowels only, we see a marked dissimilarity in the auditory bandwidth required for effective vowel classification. As stated previously the spoken back vowel will achieve a 90% vowel classification at a bandwidth of 0-6 Bark, however the sung back vowel classification requires a much larger bandwidth 0-12 Bark to achieve the same level of classification. This would be indicative of the fact that both PLP spectral peaks are required by the human ear to discriminate sung back vowels.

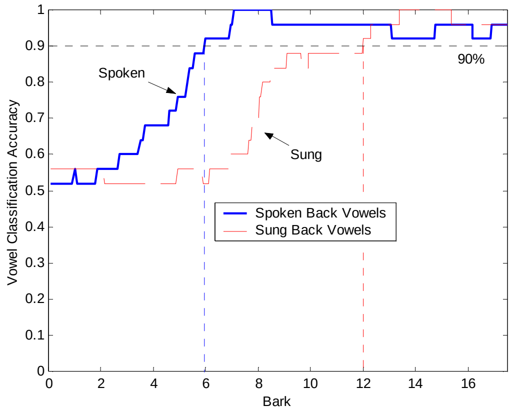

Figure 8: Vowel classification accuracy of selected auditory bandwidths for spoken and sung back vowels (/a/, /ɜ/, /ɒ/, /ɔ/, and /ʊ/).

These results indicate that the sung F_2', whilst remaining central due to the nature of the underling higher acoustic formant structure, is most likely to have a phonetic perceptual importance.

5. Conclusion

We have introduced in this paper the application of perceptual linear predication as a means of classifying the acoustic-auditory domain differences between the spoken and sung vowels. Of notable importance is the presence of a reasonable frequency stable effective second formant (F_2') across all sung vowels. This reasonably stable effective second formant is hypothesised to be the result of the underling singer's formant merging with the effective second formant.

The dimensionality of the data used for this study is small and, so therefore, no general claims can be made about the acoustic–auditory nature of the sung vowel. Discussions have been made that imply that the F_2' plays an important role in the perceptual significance of sung back vowels as a function of its spectral shape rather than its centre frequency.

6. References

Bladon, R. A. W. (1983) "Two-formant models of vowel perception: shortcomings and enhancements", *Speech Communications 2:305-313.*

Bladon, R. A. W. and Fant, G. (1978) "A two-formant model and the cardinal vowels." *STL-QPSR 1/1978. Pg 1-8.*

Bloomploft, G. & Plomp, R. (1985) "Spectral analysis of sung vowels. II. The effect of fundamental frequency on vowel spectra", *J. Acoust Soc Am 77:1580:1588.*

Chiba, T. & Kajiyama, M. (1941) *The Vowel: Its Nature and Structure* (Tokyo Kaiseikan, Tokyo).

Chistovich, L. A., Sheikin, R. L., and Lublinskaja, V. V. (1978). "Centres of gravity and spectral peaks as the determinates of vowel quality." In *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Ohman (Academic, New York). 143-157.

Clermont, F. (1992) "Formant-Contour parameterisation of vocalic sounds by temporally-constrained spectral matching." *Proc. of the 4th Australian Conference on Speech Science and Technology, Brisbane, Australia, 48-53.*

Clermont, F. (2002) "Systemic comparison of spoken and sung vowels in formant-frequency space." *Proc. 9th Australian International Conference on Speech Science and Technology. 124-129.*

Clermont, F. & Mokhtari, P. (1994) "Frequency-band specification in cepstral distance computation." *Proc. 5th Australian International Conference on Speech Science and Technology. 354-359.*

Delattre, P., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952). "An experimental study of the acoustics determinates of vowel colour" *Word **8**, 195-210.*

Fant, G. & Risberg, A. (1962) "Auditory matching of vowels with two formant synthetic sounds" *STL-QPRS 4:7-11.*

Fujimura, O. (1967). "On the second spectral peak of front vowels: a perceptual study of the role of the second and third formants," *Lang. Speech, **10**, 181-193.*

Hermansky, H. (1990) "Perceptual linear predictive (PLP) analysis of speech" *J. Acoust. Soc. Am **87**(4), April 1990.*

Ladefoged, P. (2001) *A course in phonetics*. 4th Edition. Harcourt College Publishers, Florida, USA.

Mokhatari, P. & Clermont, F. (1994) "Contributions of selected spectral regions to vowel classification accuracy." *ICSLP. 1923-1926.*

Mokhatari, P. & Clermont, F. (1996) "A methodology for investigating the vowel-speaker interactions in the acoustic-phonetic domain." *Proc. 6th Australian International Conference on Speech Science and Technology. 127-132.*

Sundberg, J. (1970) "Formant structure and articulation of spoken and sung vowels". *Folia Phoniatrica 22, 28-48.*

Sundberg, J. (1974) "Articulatory interpretation of the singing formant", *J. Acoust. Soc. Am. 55: 838-844.*

Vallabha, G. K. and Tuller, B. (2002), "Systematic errors in the formant analysis of steady-state vowels". *Speech Communication. 38, 141-160.*

Accepted after full paper review