Formant-pattern estimation guided by cepstral compatibility

Frantz Clermont[1], *Philip Harrison*[2], *and Peter French*[3]
[1,2,3]*J.P. French Associates & University of York, UK*
{frantz.clermont|pth|jpf}@jpfrench.com

The following questions arise in dealing with the problem of estimating vocal-tract, resonance-patterns (or F-patterns) that best explain the observed spectrum: (a) how "*robust*" is one F-pattern estimator *by comparison* with another?; (b) how "*reliable*" is an F-pattern obtained from a *given* estimator? A recent attempt (Harrison, 2004) at elucidating the former question has yielded a clear perspective of caution, owing to the lack of consistency uncovered amongst a number of F-pattern estimators. Here we investigate the latter question of reliability using a method that is based on the linear-prediction (LP) model (Markel & Gray, 1976) of speech production.

The method combines the analysis-by-synthesis procedure pioneered by Bell et al. (1961), with spectral-shape properties of a mainstay in speech signal parameterisation: (1) the *low-order cepstral sequence* obtained through LP analysis carries the bulk of resonance information about the vocal tract; (2) the *un-weighted* Euclidean distance between a paired set of cepstral sequences is equivalent to a spectral shape distance (Gray & Markel, 1976); (3) the *index-weighted* form of the cepstral distance affords an added sensitivity around peak locations (Yegnanarayana & Reddy, 1979). Thus, the problem of objectively selecting the most compatible F-pattern with respect to the observed spectrum can be recast entirely in cepstral terms without having to explicitly generate spectral shapes. By converting a set of candidate F-patterns into *simplified cepstra* (Clermont, 1991), their relative compatibility with the observed cepstrum becomes quantifiable by means of the peak-sensitive, cesptral distance. The minimum of the resulting set of distances is then used to select the simplified cepstrum that provides the most compatible F-pattern.

The results thus far are based on a small dataset of LP-cepstra measured through the steady-state intervals of three categories ([FLEECE, PALM, SCHWA] after Wells (1982)) of British English vowels in CV(C) contexts, which were recorded over the telephone by one adult-male, native speaker. The compatible F-patterns are found to be closely aligned with the fact that the index-weighted, cepstral distance tends to block spurious peaks, even those that may have narrow bandwidths but nevertheless contribute relatively little to the overall spectral match. This behaviour seems unaffected by the fluctuations in bandwidths expected from LP analysis and, thus, the minimum cepstral distance gains in stability as a criterion for selecting the pattern of peaks that are best matched to the observed spectral shape. These preliminary findings are interpreted as supportive evidence for the method of cepstral compatibility introduced in this paper.

References

Bell, C. G., H. Fujisaki, J.M. Heinz, K.N. Stevens, and A.S. House (1961). Reduction of speech spectra by analysis-by-synthesis techniques. *J. Acoust. Soc. Am.,* **33**, 1725-1736.

Clermont, F. (1991). *Formant-Contour Models of Diphthongs: A Study in Acoustic Phonetics and Computer Modelling of Speech*. PhD Thesis: Australian National University.

Gray, A. H. and J.D. Markel (1976). Distance measures for speech processing. *IEEE Trans. Acoustics, Speech and Signal Processing*, **24**, 380-391.

Harrison, P.T. (2004). *Variability of Formant Measurements*. MA Thesis: University of York.

Markel, J.D. and A.H. Gray (1976). *Linear Prediction of Speech*. Springer Verlag.

Wells, J.C. (1982). *Accents of English.* Cambridge University Press.

Yegnanarayana, B. and R. Reddy (1979). A distance measure based on the derivative of linear prediction phase spectrum. *Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing,* 744-747.