Acoustic Evaluation of Sources of Vibrato: Harmonic Analysis of the Singing Voice

Satoshi Imaizumi, Haruhito Saida, Yoko Shimura
and Hajime Hirose

Introduction

The sources of vibrato in the singing voice can be investigated by measuring[1)] a) the neural activities in the central nervous system; b) the muscle action potential (EMG)[2-4)]; c) the respiratory, laryngeal and articulatory movements caused by the muscle activity[5,6)]; d) the variations of airflow and pressure within the vocal tract caused by the various movements[7)]; e) the acoustic properties of the resultant radiated sound, such as variations in periodicity and amplitude[8-14)]; and f) the properties of the perceived vibrato[15,16)].

EMG measurement, for instance, is invasive and not always feasible with professional singers. Some measurements of the respiratory, phonatory and articulatory movements may affect singing style. Among the above-mentioned possibilities, therefore, an acoustic approach must be the most convenient if it can provide fruitful insights on the sources of vibrato. Although many acoustic analysis procedures have been reported, no one acoustic approach has been accepted as an effective method for obtaining the sources of vibrato, one of the reasons being the complicated relations between the subglottal pressure, the fundamental frequency and the formant frequencies[7,14)].

This paper describes a novel method of investigating the acoustic characteristics of vibrato by looking at the relationships between the fundamental frequency, the formant frequencies and the amplitude of the singing voice.

Method

Working Hypothesis

Four factors contributing to the acoustic characteristics of vibrato -- the subglottal pressure, the fundamental frequency, the voice source spectrum and the formant frequencies -- were taken into account here. The contribution of each factor was hypothesized to be as follows.

Case A: If the main factor determining vibrato is a change in only the fundamental frequency, keeping the other factors relatively constant, then the overall sound pressure level variations could conceivably be caused by variations in the fundamental frequency (F_0), as these variations in F_0 cause the higher partial tones to move toward and away from the formants. In Case A, therefore, partial tones or harmonics move along the vocal tract transfer function $T(f)$ as shown in Figure 1 (a) when F_0 moves up and down.

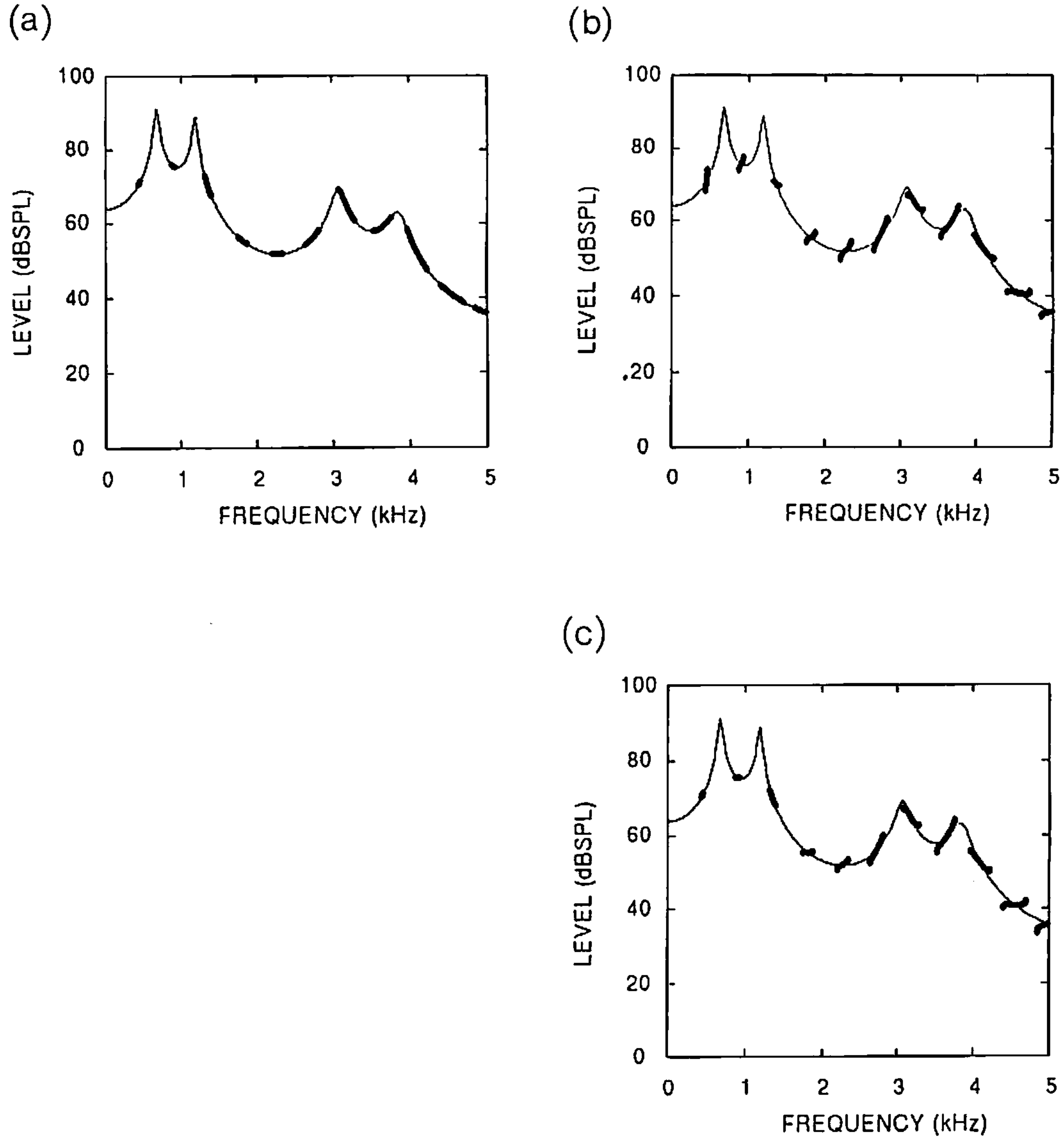

Figure 1. Hypothesized relationships between formant structure *T(f)* (solid line), fundamental frequency (F_0) and level of harmonics (circles). (a) Case A: Only the contribution of variations in F_0 is significant, (b) Case B: The subglottal pressure contributes significantly, (c) Case C: The voice source spectrum changes and synchronized to the vibrato.

Furthermore, the phase relationships among the level variations of the harmonics can be complex depending on their proximity to the formants. Some co-vary in phase, while others out of phase. For instance, in Figure 1(a), when F_0 increases, the level of the first harmonic increases, and the level of the third decreases.

Case B: If the contribution of variations in the subglottal pressure is large, in addition to the contribution of the F_0 variation, the overall sound pressure level could conceivably be caused by variations in both factors. When the subglottal pressure increases, F_0 usually increases, and the level of harmonics also increases. Therefore the harmonics vary across the vocal tract transfer function $T(f)$, or traverse it as shown in Figure 1 (b). Furthermore, the harmonic variations tend to be in phase.

Case C: If not only the variations in F_0 but also those in the voice source spectrum significantly contribute to vibrato, it may be observed that lower harmonics vary along $T(f)$, while the higher harmonics traverse $T(f)$ as shown in Figure 1 (c). This is to be expected because the voice-source spectrum is supposed to be flatter for higher F_0 than for lower F_0 in the singing voice.

By analyzing the frequency-level relationships of the harmonics as described below, it may be possible to estimate the formants more accurately than by conventional methods such as a linear predictive method.

Parameter Estimation

A new method for estimating formants applicable even for high pitched singing voices with vibrato is proposed in this section. This method uses vibrato to improve the estimation reliability.

Singing voice samples simultaneously recorded with electroglottogram (EGG) from one soprano and one mezzo-soprano were analyzed. The voice and EGG signals were digitized through a 12-bit A/D converter at a sampling rate of 40 kHz for each signal.

At first, using a peak picking method[17,18], the local maximum points of voice signal, s(sample number:j), which could correspond to vocal excitation epochs were detected successively. Here, we write $L(i)$ for the i-th pitch location, $p(i)=L(i)-L(i-1)$ for the i-th pitch period, $F_0(i)=1/p(i)$ for the i-th fundamental frequency; and $e(i)$ for the amplitude at $L(i)$; for $i=0, 1, ..., I$, where I was the total number of pitch periods extracted. $SPH(i)$ was defined as the peak-to-peak value of $s(j)$ for each i-th pitch period.

The level of the n-th harmonic for the i-th pitch period, $H_n(i)$, was calculated as follows.

$$H_n(i)=10.0\log_{10}(h_{rn}(i)^2+h_{in}(i)^2)$$

$$h_{rn}(i)=\sum_{j=L(i-1)}^{L(i)-1} s(j)\cos(6.28jn/p(i)) \qquad (1)$$

$$h_{in}(i)=\sum_{j=L(i-1)}^{L(i)-1} s(j)\sin(6.28jn/p(i))$$

The n-th formant frequency, F_n, and its bandwidth, B_n, were estimated by fitting the measured harmonics for one cycle of vibrato with the product of three factors *P(f)*, *V(f)* and *R(f)*. *F(f)* was the product of the four lower formants, *S(f)* the voice source spectrum and *R(f)* the higher pole compensation factor. These factors were defined as follows,

$$V(f)=A_v(F_v^2+B_v^2)/(T_pT_m)^{1/2}$$

$$T_p=(f+F_v)^2+B_v^2 \qquad (2)$$

$$T_m=(f-F_v)^2+B_v^2$$

where Fv is the adjustable pole frequency and Bv is the adjustable bandwidth representing the voice source spectrum. A_v is the adjustable gain factor representing the voice source magnitude.

$$P(f)=\prod_{k=1}^{4} M_n(f)$$

$$M_n(f)=A_v(F_n^2+B_n^2)/(T_pT_m)^{1/2}$$

$$T_p=(f+F_n)^2+B_n^2 \qquad (3)$$

$$T_m=(f-F_n)^2+B_n^2$$

$$R(f)=0.54(f/F_1)^2+0.00143(f/F_1)^4+20.0\log_{10}(f/100) \qquad (4)$$

The last term of *R(f)* represents the radiation characteristics.

To estimate the n-th formant frequency, F_n, and its bandwidth, B_n, a least square error method was used, in which parameters were determined so as to minimize Q.

$$Q = \sum_{n=0}^{N(i)} \sum_{i=0}^{I} (H_n(i) - T(nF_0(i)))^2 \qquad (4)$$

where $N(i)$ is the number of harmonics below 5kHz for the i-th pitch period.

To estimate the degree of abduction or adduction of the glottis[5)], *EGGOQ(i)* and *EGGH(i)* were estimated as shown in Figure 2. In this figure, "*A*" was defined as the duration between the local maximum and the local minimum peaks of the differentiated EGG in a pitch period, and "*B*" the duration between a local maximum peak and the next local maximum peak. *EGGOQ* was defined as *B/(A+B)*. Since *EGGOQ* tends to increase when the duration of the opened glottis in a pitch period increases, thus it is possible to consider that it represents the ratio of the duration of the opened glottis to the pitch period. The *EGGH* was defined as the peak-to-peak amplitude of the EGG in a pitch period. The *EGGH* was measured, because it is possible to consider that a greater amplitude in the EGG signal represents less electrical impedance through the larynx and a greater contact area between the vocal folds.

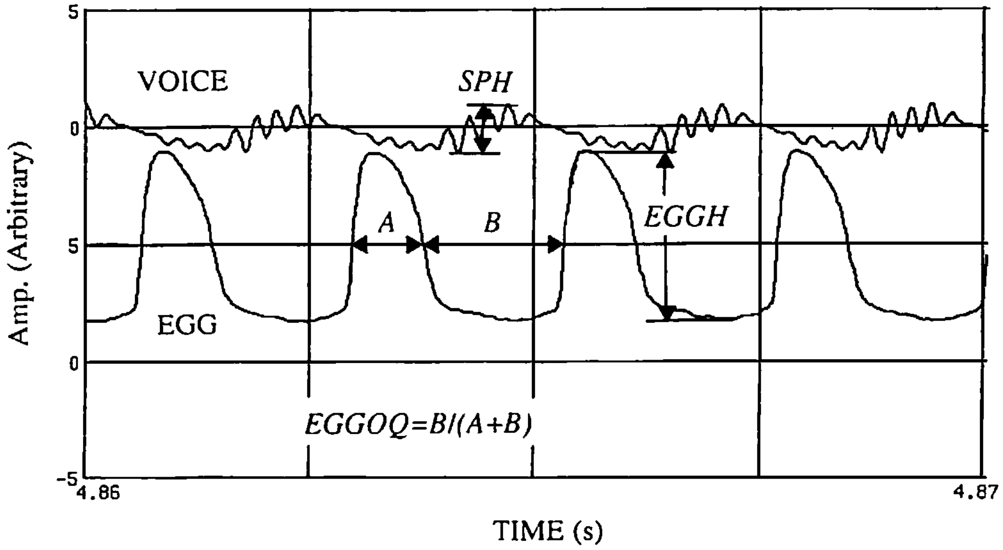

Figure 2. Definitions of *EGGOQ* and *EGGH*. "*A*" was defined as the duration between the local maximum and local minimum peaks of the differentiated EGG in a pitch period, and "*B*" the duration between a local maximum peak and the next local maximum peak. The *EGGOQ* was defined as *B/(A+B)*. The *EGGH* was defined as a peak-to-peak amplitude of the EGG in a pitch period.

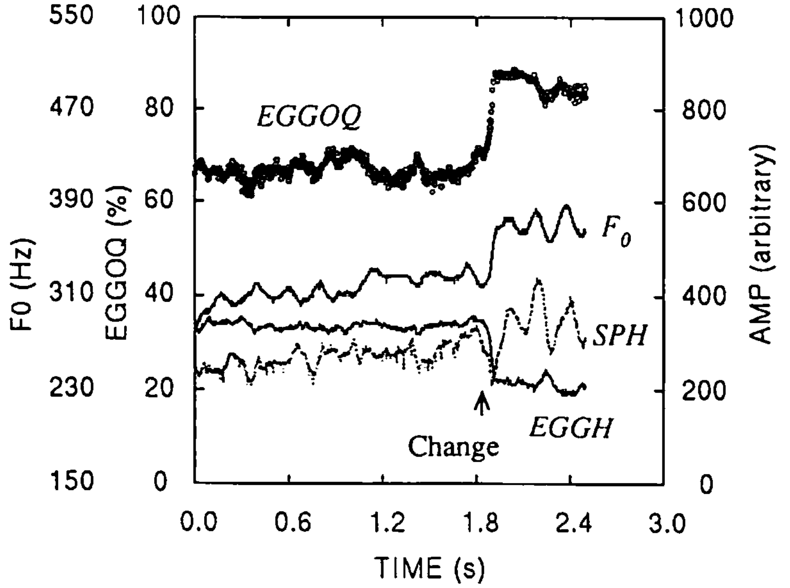

Figure 3. The results of the EGG analysis for the vowel /a/ produced with ascending pitch from c^1 to c^2. The point "Change" indicates the point where the singer changed voice register from "Chest" to "Mid."

Results and Discussion

Figure 3 shows the results of the EGG analysis for the voice sample sung by the soprano singer. She sang the vowel /a/ in an ascending pitch from c^1 to c^2. The point "Change" indicates the point where the singer changed her voice register from "Chest" to "Mid". Before this point, during "Chest" register, the *EGGOQ* was around 63%, after which it jumped to 85% to produce the "Mid" register voice. The *SPH* increased at this point, but the *EGGH* decreased from 350 to 200.

Figures 4 (a) and (b) show the results of the formant estimation and the harmonic extraction from the soprano voice whose waveform is shown in Figure 4 (b). The analyzed voice sample has an average fundamental frequency of 458Hz, and it has a 5.5Hz vibrato.

As shown in Figure 4 (a), the level of the lower harmonics numbered 1 to 3 fit the estimated average frequency characteristics *T(f)*, while the harmonics higher than the 3rd did not fit *T(f)*, but rather traversed it.

Furthermoer, as shown in Figure 4 (b), the phase relationships among the level

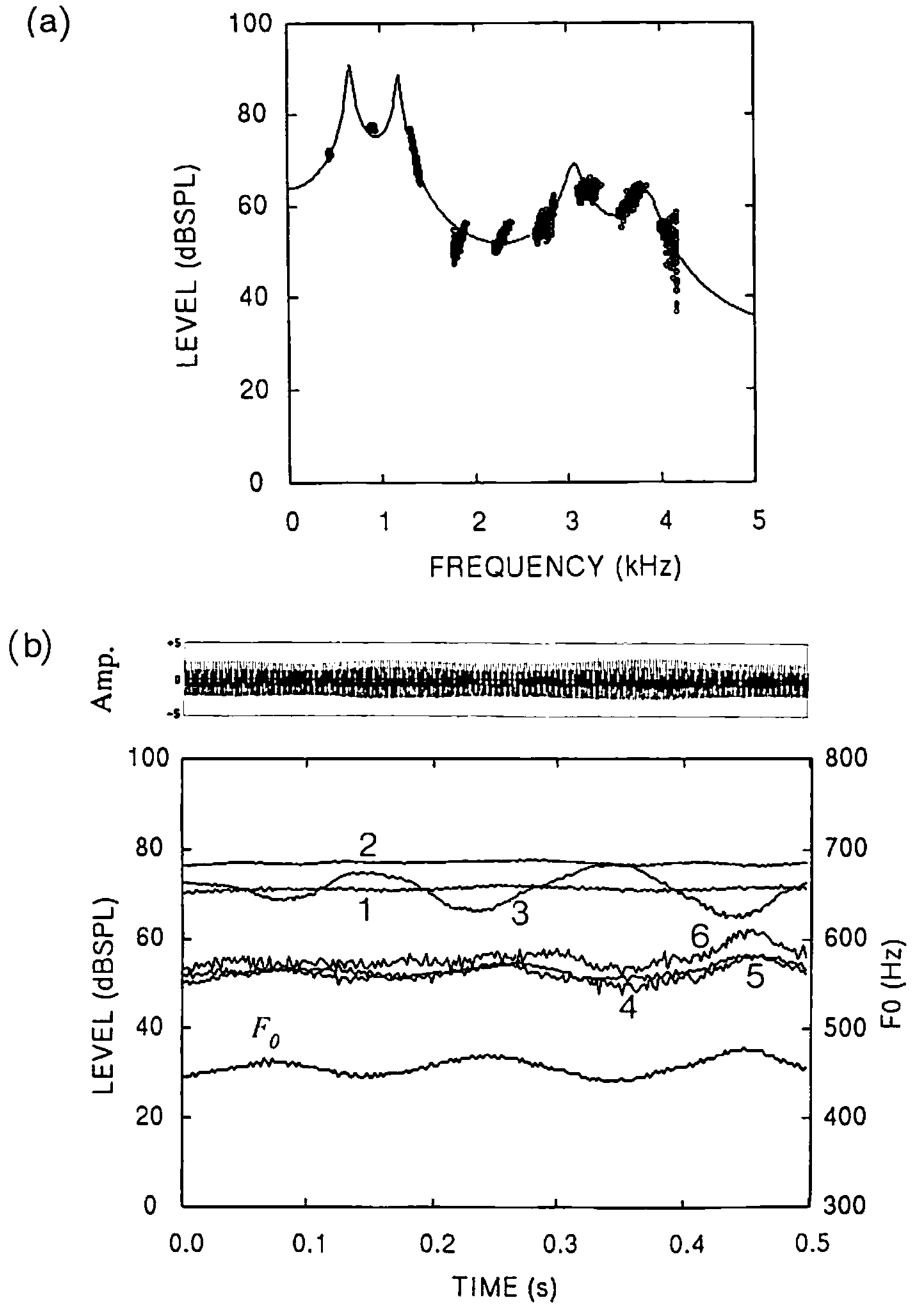

Figure 4. (a) The results of the formant estimation, (b) the phase relationships among the level variations of the harmonics and F_0, and the waveform of the soprano voice analyzed.

variations of the harmonics are somewhat complex depending on their proximity to the formants. The level of the 3rd harmonic varys out of phase with F_0 and the other harmonics. The harmonics except the 3rd co-vary in phase with F_0. The variation in the overall amplitude of this soprano voice is mainly determined by the level variation of the 3rd harmonic.

These results shown in Figures 4 indicate that this soprano vibrato sample was controlled as predicted by the hypothesized Case C. In this case, not only the variations in F_0 but also those in the voice source spectrum contributed to the vibrato. The lower harmonics vary along *T(f)*, while the higher harmonics traverse *T(f)* in as predicted in Figure 1 (c).

Variations in the voice source spectrum can be caused by proper laryngeal adjustments or by control of the subglottal pressure. The EGG results shown in Figure 3 indicate that the *EGGOQ* was relatively stable compared to the drastic change observed when the singer changed her vocal register from "Chest" to "Mid", but it still fluctuated slightly in such a way that it decreased when F_0 increased. This variation may indicate that the open quotient of the glottis decreased when F_0 increased during the vibrato, which may further suggest that the vocal folds were stretched/adducted slightly when F_0 increased. In addition to this speculation, the fact that the level of the harmonics lower than the 3rd remained stable indicates that, at least in this case, laryngeal adjustment may play a more significant role in vibrato than subglottal pressure.

This paper reports only examples of Case C, but so far there is no reason to reject Cases A and B from our experiments. Obviously, we need more detailed analyses for a large number of singers. Further results on formant estimation and vibrato analyses for sung vowels will be reported in the future.

Conclusions

In order to investigate the sources of vibrato in the singing voice through an acoustic approach, a new acoustic analysis method was proposed, which utilizes the level-frequency variability of harmonics and is applicable even for high-pitched singing voices with extensive vibrato. Using this method, the interrelationships between the fundamental frequency, the level of the harmonics, the formant frequencies and the overall amplitude of the singing voice were analyzed. The results showed that the lower harmonics varied along the estimated average vocal tract transfer function including the voice source characteristics, while the higher harmonics intersect it. It was suggested that not only the variations in F_0 but also those in the voice source spectrum contribute to vibrato, and that laryngeal adjustment may play a more significant role in vibrato than subglottal pressure at least for the singer analyzed here.

Acknowledgement

This study was supported by a Grant-in-Aid for Scientific Research on Priority Areas, the Ministry of Education, Science and Culture, Japan, "Emotional Information" (No.04236102)

References

1) M. Rothenberg, D. Miller and R. Molitor:" Aerodynamic investigation of sources of vibrato", Folia Phoniatr., 40, 244-260 (1988).
2) M. Hirano, W. Vennard and J. Ohara: "Regulation of register, pitch and intensity of voice," Folia Phoniatr., 22, 1-20 (1970).
3) S. Niimi, S. Horiguchi, and N. Kobayashi:"The physiological role of the sternothyroid muscle in phonation: An electromyographic observation," Ann. Bull. RILP, 22, 163-170 (1988).
4) T. Shipp, E. T. Doherty and S. Haglund: "Physiologic factors in vocal vibrato production," J. Voice, 4(4), 300-304 (1990).
5) L. R. Brown and R. C. Sherer:"Laryngeal adduction in trillo,"J. Voice, 6(1), 27-35 (1992).
6) D. H. Hicks and E. Teas:" An electroglottographic study of vocal vibrato," J. Voice, 1(2), 142-147 (1987).
7) J. Sundberg: "What's so special about singers?," J. Voice, 4(2), 107-119 (1990).
8) Y. Horii: "Acoustic analysis of vocal vibrato: A theoretical interpretation of data," J. Voice, 3(1), 36-43 (1989).
9) H. K. Schutte and D. G. Miller: "Acoustic details of vibrato in tenor high notes," J. Voice, 5(3), 217-223 (1991).
10) H. B. Rothman and A. A. Arroyo:"Acoustic variability in vibrato and its perceptual significance," J. Voice, 1(2), 123-141 (1987).
11) J. Hakes, T. Shipp and E. T. Doherty: " Acoustic properties of straight tone, vibrato, trill, and trillo," J. Voice, 1(2), 148-156 (1987).
12) D. Myers and J. Michel: "Vibrato and pitch transitions," J. Voice, 1(2), 157-161 (1987).
13) L. A. Ramig and T. Shipp: "Comparative measures of vocal tremor and vocal vibrato," J. Voice, 1(2), 162-167 (1987).
14) R. F. Coleman, J. Hakes, D. M. Hicks, J. F. Michel, L. A. Ramig and H. B. Rothman: "Discussion on vibrato," J. Voice, 1(2), 168-171 (1987).
15) C. E. Seashore: Psychology of music (Dover Publications, Inc., New York, 1967).
16) M. S. Benolken and C. E. Swanson:" The effect of pitch-related changes on the perception of sung vowels," J. Acoust. Soc. Amer., 87(4), 1781-1785 (1990).
17) S. Imaizumi:"Acoustic measures of roughness in pathological voice," J. Phonetics, 14, 457-462 (1986).

18) S. Imaizumi:"Acoustic measurement of pathological voice qualities for medical purposes," Proc. of IEEE-IECEJ-ASJ International Conf. on Acoustics. Speech. and Signal Processing, ICASSP 86 (Tokyo), 677-680 (1986).