Linear Transformation from Full-Band to Sub-Band Cepstrum

Frantz Clermont

School of Culture, History and Language, Australian National University, Canberra, Australia
& Forensic Speech and Acoustics Laboratory, J.P. French Associates, York, England
dr.fclermont@gmail.com

Abstract

This paper demonstrates the possibility of estimating the cepstrum for a sub-band region of the full-band spectrum by a linear transformation of the full-band cepstrum. The parametric formulation of the transformation allows the selection of any sub-band within the full-band's frequency range. The result of the transformation is a Fourier series of band-limited cepstral coefficients (BLCCs) representing the selected sub-band. In practice, the upper bound of the BLCC series may be fixed at $(M \times W)$ without significant loss in spectral resolution, where M is the number of full-band coefficients, and W is the fraction of the full-band spectrum occupied by the sub-band's width.

Index Terms: cepstrum, full band, sub-band, Fourier series, linear transformation

1. Introduction

Sub-band cepstra are commonly estimated using the filter-bank method [1,2]. The bandpass filters are designed with specific widths, and their locations along the frequency axis are also pre-determined. While these constraints have not impeded progress in speech and speaker classification, they restrict the ability to explore the full potential of sub-bands.

Here we propose an alternative method, which bypasses the problem of re-adjusting the filter-bank design and obviates the cost of analysing the speech signal for every sub-band of interest. Our method requires only the full-band cepstrum and a linear transformation to contain it within the limits of any sub-band.

Sections 2 and 3 develop the formulae to transform full-band into band-limited cepstral coefficients (BLCCs). Section 4 illustrates these for two sub-bands selected from speech cepstra. The key features of our band-limiting method are summarised in Section 5 with a brief overview of application possibilities.

2. Full-band cepstrum

The log-magnitude spectrum $S(\omega)$ can be defined as a Fourier cosine series of the so-called cepstral coefficients C_k:

$$S(\omega) = \sum_{k=1}^{M} C_k \cos(k\omega)\,, \quad 0 \le \omega \le \pi \qquad (1)$$

Truncating this series after M terms yields a cepstrally-smoothed representation of $S(\omega)$ across the entire available bandwidth. The C_k are thus interpreted as full-band coefficients.

3. Band-limited cepstrum (BLC)

3.1. Band-limiting parameters

Let ω_1 and ω_2 be the limits of a sub-band, and let the associated change of variables $\omega \rightarrow \omega'$ be as follows:

$$\omega' = \pi \left[\frac{(\omega-\omega_1)}{(\omega_2-\omega_1)}\right], \quad \omega_1 \le \omega \le \omega_2 \qquad (2)$$

From Eq. (2) it is easy to express ω as a function of ω':

$$\omega = \omega_1 + \left[\frac{(\omega_2-\omega_1)}{\pi}\right]\omega' = \omega_1 + W\omega' \qquad (3)$$

where $0 \le \omega' \le \pi$, and where the scalar W is the ratio of the sub-band's width to the full-band's frequency range:

$$W = \left[\frac{(\omega_2-\omega_1)}{\pi}\right], \quad 0 < \mathrm{W} \le 1 \qquad (4)$$

3.2. BLC formulation

The band-limited analogue of Eq. (1) is defined in Eq. (5) as a Fourier cosine series representing the spectral region of the full band delimited by the sub-band:

$$S(\omega(\omega')) = C'_0 + \sum_{l=1}^{N} C'_l \cos(l\omega')\,, \quad 0 \le \omega' \le \pi \qquad (5)$$

The notation $\omega(\omega')$ indicates that the argument ω is a band-dependent function of ω' via Eq. (3), C'_l is the l-th band-limited cepstral coefficient (BLCC), and N is the upper bound of the BLCC series. The zeroth-order coefficient retains the level of the band-limited spectral region, and the other coefficients account for the spectral shape. While N theoretically goes to ∞, a much lower bound is sufficient to preserve the overall shape. This is discussed in Section 3.4 and illustrated in Section 4.

3.3. BLCC derivation

3.3.1. Fourier cosine coefficients: Formulae & solutions

The cepstral coefficients C'_l in Eq. (5) are derived from the standard Fourier integration formulae, as follows:

$$C'_{l=0} = \frac{1}{\pi}\int_0^{\pi} S(\omega(\omega'))\, d\omega' \qquad (6)$$

$$C'_{l>0} = \frac{2}{\pi}\int_0^{\pi} S(\omega(\omega'))\cos(l\omega')\, d\omega' \qquad (7)$$

$S(\omega(\omega'))$ is next replaced with the cosine series in Eq. (1) and ω with the right-hand side of Eq. (3) to yield:

$$C'_{l=0} = \frac{1}{\pi}\int_0^{\pi}\left\{\sum_{k=1}^{M} C_k \, cos[k(\omega_1 + W\omega')]\right\} d\omega' \qquad (8)$$

$$C'_{l>0} = \frac{2}{\pi}\int_0^{\pi}\left\{\sum_{k=1}^{M} C_k \, cos[k(\omega_1 + W\omega')]\right\} \cos(l\omega')\, d\omega' \qquad (9)$$

The finite sum over cepstral coefficients in Eqs (8) and (9) justifies reversing the order of integration and summation. The former becomes a separate operation, the result of which is inserted into the sum as a weighting coefficient. This is reflected in Eq. (10), where the band-limited coefficients C'_l are expressed as a weighted sum of the full-band coefficients C_k:

Table of Contents for this manuscript

$$C'_l = \sum_{k=1}^{M} a_{lk} \cdot C_k, \quad l = 0,1,\ldots,N \tag{10}$$

The formulae for a_{lk} are expressed below as functions of ω_1 and ω_2. There are 3 cases depending on certain values of l:

if $l > 0$ and $l \neq kW$,

$$a_{lk} = \frac{2(kW)}{\pi[l^2-(kW)^2]}[(-1)^{l+1}\sin(k\omega_2) + \sin(k\omega_1)] \tag{11a}$$

if $l > 0$ and $l = kW$,

$$a_{lk} = \cos(k\omega_1) \tag{11b}$$

if $l = 0$,

$$a_{lk} = \frac{1}{k(\omega_2-\omega_1)}[\sin(k\omega_2) - \sin(k\omega_1)] \tag{11c}$$

Equation (11a) follows from evaluating the integral in Eq. (9) with adaptations of trigonometric solutions given in [3]. Equation (11b) is simply the limit of Eq. (11a) as $l \rightarrow \text{kW}$. Equation (11c) also flows from Eq. (11a) by substituting l for 0 and dividing the result by 2. To decide which formula to use for $l > 0$, it is numerically desirable to test the condition $|(l - kW)| < \varepsilon$, where ε is a small positive number.

3.3.2. *The weighted sum in matrix form:* $\mathbf{c}' = \mathbf{Ac}$

The weighted sum in Eq. (10) implies a *linear transformation* from C_k to C'_l. Let **c** be a column vector $(M \times 1)$ of C_k, $\mathbf{c}'$ a column vector $((N+1) \times 1)$ of C'_l, and **A** the transformation matrix $((N+1) \times M)$ with elements a_{lk}. Equation (10) can thus be recast in the matrix form $\mathbf{c}' = \mathbf{Ac}$ laid out below. This is not only a convenient way of encapsulating Eqs (10) and (11), but one that is also useful for computer implementation.

$$\begin{bmatrix} C'_0 \\ C'_1 \\ \vdots \\ C'_l \\ \vdots \\ C'_N \end{bmatrix} = \begin{bmatrix} a_{0,1} & \cdots & a_{0,k} & \cdots & a_{0,M} \\ a_{1,1} & \cdots & a_{1,k} & \cdots & a_{1,M} \\ \vdots & & \vdots & & \vdots \\ a_{l,1} & \cdots & a_{l,k} & \cdots & a_{l,M} \\ \vdots & & \vdots & & \vdots \\ a_{N,1} & \cdots & a_{N,k} & \cdots & a_{N,M} \end{bmatrix} \begin{bmatrix} C_1 \\ \vdots \\ C_k \\ \vdots \\ C_M \end{bmatrix} \tag{12}$$

3.4. A practical bound for truncating BLCC series

Recall that Eq. (5) is a Fourier series representation of $S(\omega)$ over the interval $[\omega_1, \omega_2]$, just as Eq. (1) is one for the same function but over $[0, \pi]$. As a finite sum of sinusoids, $S(\omega)$ and its derivatives are continuous over $[0, \pi]$ and hence also over $[\omega_1, \omega_2]$. This guarantees that the series in Eq. (5) will converge uniformly for increasing values of the upper bound N. A relevant question is how large N needs to be in practice.

Our reasoning for truncating the BLCC series is as follows. If M cepstral coefficients represent the full-band spectrum with a certain resolution, then roughly the same resolution for the sub-band region should be achievable with $N = (M \times W)$, hereafter referred to as MW. This means retaining from M the fraction W of the full band occupied by the sub-band's width. Note that MW will generally not be an integer, and hence it must be rounded to be useable as a coefficient index.

Section 4.1 illustrates the BLCC series for two selected sub-bands, one narrower than the other. Section 4.2 shows the corresponding spectral fits around $N = MW$. In Section 4.3, a formula for estimating the truncation error is derived and applied to the same sub-bands.

4. Numerical illustrations

4.1. A glimpse at BLCC series for two sub-bands

Table 1 lists three sets of cepstral coefficients which were extracted at the frame marked vertically in Fig. 1. The full-band C_k (order M=14) were obtained by discrete-cosine transform of the log-magnitude FFT spectrum ranging from 0 to 5 kHz.

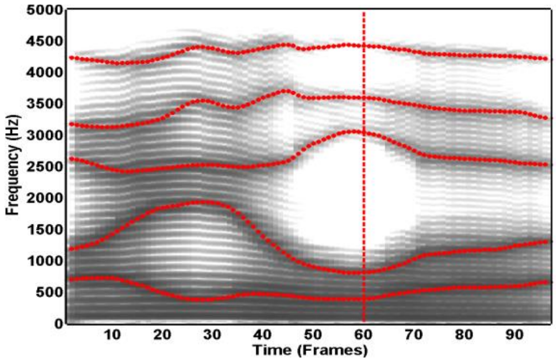

Figure 1: *Spectrogram of "IOWA" with superimposed formant-frequency tracks. A vertical line is marked at the frame of interest in this section and in Section 4.2.*

Equations (11) and (12) were used to generate BLCCs for these sub-bands: [0.1,0.9]-kHz and [2.5,4.0]-kHz. For the sake of illustration, the upper bound N of the BLCC series was set to 14, thus matching the number M of full-band coefficients and extending the BLCC series well beyond $MW = 2$ for the narrower sub-band and $MW = 4$ for the wider one.

Table 1. C*epstral coefficients obtained at the frame marked in Fig. 1. The BLCCs* C'_l *for the two selected sub-bands are highlighted up to their respective* MW*.*

Coeff. index	Full-band C_k [0.0,5.0] kHz	BLCCs C'_l [0.1,0.9] kHz	BLCCs C'_l [2.5,4.0] kHz
0		2.5281	0.0865
1	0.9136	-0.1700	-0.2540
2	0.9127	-0.5929	-0.6726
3	1.3018	0.0108	-0.0979
4	-0.1548	-0.0898	-0.1701
5	-0.3611	0.0136	-0.0119
6	-0.2508	-0.0385	-0.0509
7	-0.2258	0.0077	-0.0087
8	-0.4771	-0.0214	-0.0253
9	0.0030	0.0048	-0.0054
10	-0.0895	-0.0136	-0.0155
11	0.1266	0.0032	-0.0036
12	-0.1663	-0.0094	-0.0106
13	-0.0302	0.0023	-0.0026
14	-0.0555	-0.0069	-0.0077

The zeroth-order C'_0 is much larger in the [0.1,0.9]-kHz range, which indicates a prominent region in the lower part of the spectrum. The next C'_l exhibit a consistent trend for both sub-bands: A drop in magnitude is noticeable after MW, with a subsequent decay of the BLCC series towards zero. This occurs 2 coefficients later for the sub-band [2.5,4.0]-kHz, whose width is about twice that of the narrower sub-band.

4.2. Fitting sub-band spectra with BLCCs

Is the proposed truncation after $N = MW$ detrimental to the spectral resolution in a sub-band region? To gain some insights into this question, full-band (in blue) and band-limited (in red), cepstrally-smoothed spectra are overlaid in Figs 2 and 3. These are based on the same C_k and C_l' listed in Table 1.

There is a major improvement in the spectral fit as N increases from 1 to 2 for both sub-bands. The fit then becomes very tight when $N = MW$, except at the edges where the zero slopes are likely to cause slow convergence of the BLCC series beyond MW. This may be inconsequential in practice, especially since only minor improvements are observed after MW.

In short, it could be said that the post-MW BLCCs contribute relatively little to the band-limited spectral shape. This is consistent with their decaying magnitudes noticed in Table 1.

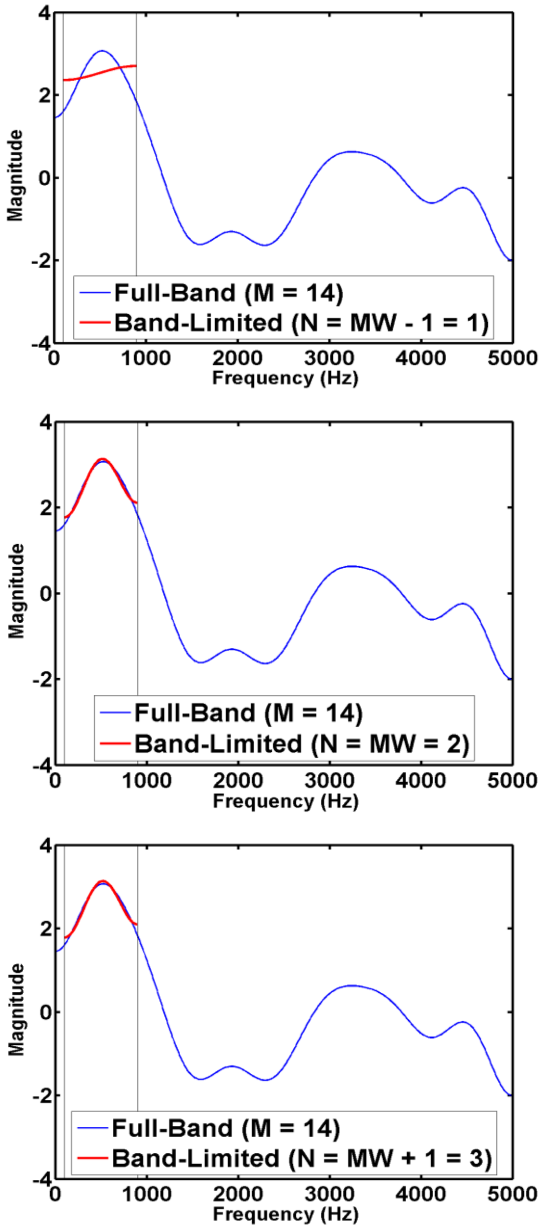

Figure 2: *Full-band & sub-band spectra based on* C_k*: [0.0,5.0]-kHz and* C_l'*: [0.1,0.9]-kHz from Table 1.*

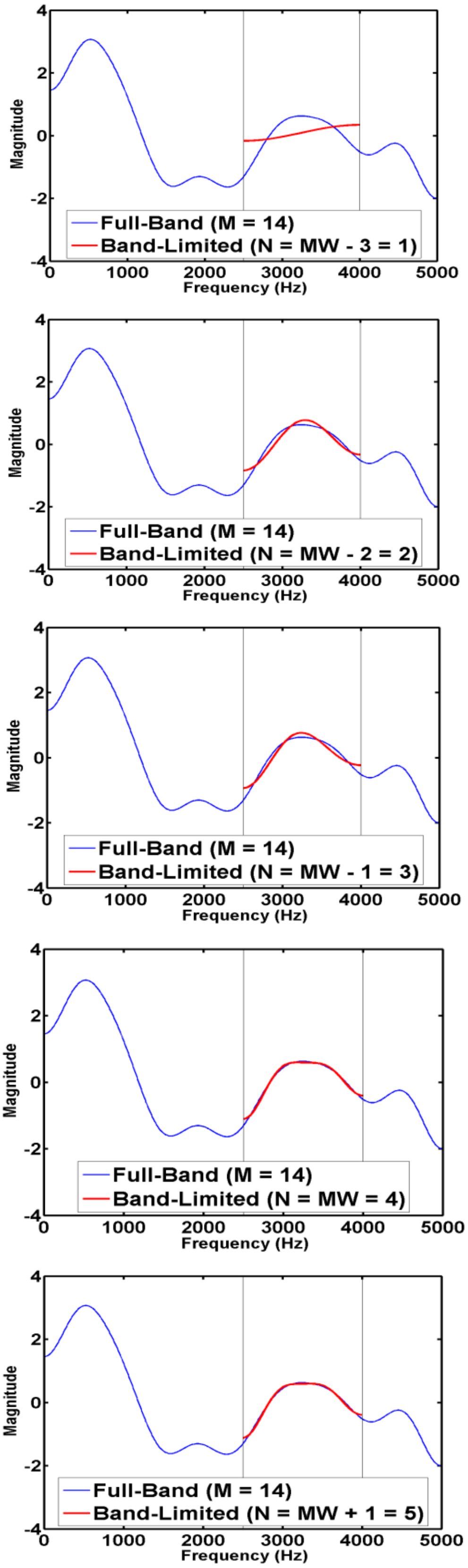

Figure 3: *Full-band & sub-band spectra based on* C_k*: [0.0,5.0]-kHz and* C_l'*: [2.5,4.0]-kHz from Table 1.*

4.3. Truncation error estimation

The mean square of the Fourier cosine expansion of the BLCC series in Eq. (5) is exploited below to observe the truncation error before and after MW. The mean square formula for BLCCs is derived in Section 4.3.1 and an estimate of the error based on this formula is proposed in Section 4.3.2. In Section 4.3.3, the latter is applied to the "IOWA" example data.

4.3.1. Mean square over a sub-band

The mean square of $S(\omega(\omega'))$ is defined as:

$$\bar{S}^2 = \frac{1}{(\omega_2-\omega_1)}\int_{\omega_1}^{\omega_2}[S(\omega(\omega'))\,]^2 d\omega \qquad (13)$$

Recall Eq. (3) to express $d\omega$ as a function of $d\omega'$:

$$\omega = \omega_1 + W\omega', \quad 0 \le \omega' \le \pi,\ \omega_1 \le \omega \le \omega_2 \qquad (14a)$$

$$d\omega = W d\omega' = \left[\frac{(\omega_2-\omega_1)}{\pi}\right] d\omega' \qquad (14b)$$

Substitute Eq. (14b) into Eq. (13), change the limits of the integral accordingly, and simplify the integrand:

$$\bar{S}^2 = \frac{1}{(\omega_2-\omega_1)}\int_{0}^{\pi}[S(\omega(\omega'))]^2 \left[\frac{(\omega_2-\omega_1)}{\pi}\right] d\omega'$$

$$= \frac{1}{\pi}\int_{0}^{\pi}[S(\omega(\omega'))]^2 d\omega' \qquad (15)$$

Replace $S(\omega(\omega'))$ with its cosine expansion from Eq. (5) to obtain:

$$\bar{S}^2 = \frac{1}{\pi}\int_{0}^{\pi}[C_0' + \textstyle\sum_{l=1}^{N} C_l' \cos(l\omega')]^2\, d\omega' \qquad (16)$$

Parseval's theorem applied to Eq. (16) yields the mean-square solution for BLCCs:

$$(\bar{S}^2)_N = (C_0')^2 + \frac{1}{2}\textstyle\sum_{l=1}^{N}(C_l')^2 \qquad (17)$$

4.3.2. Error measure

As shown in Table 1, the BLCC series tends towards zero from MW up to the upper bound M selected for illustrative purposes. One way of quantifying this behaviour is to calculate the mean-square differences between the BLCC series extended as far as M and the same series truncated one cepstral coefficient at a time. These differences give an estimate of the truncation error E, and the formula adopted derives from Eq. (17) as follows:

$$E_l = (\bar{S}^2)_M - \textstyle\sum_{l=0}^{M}(\bar{S}^2)_l \qquad (18)$$

4.3.3. Truncation error profiles for the two selected sub-bands

The BLCC analysis of the utterance "IOWA" was applied to all 97 frames of 25-msec duration (with 5-msec step size). The error measure E given in Eq. (18) was calculated at all frames and for the same sub-bands studied earlier.

Figures 4 and 5 display the (97-frame) means and standard deviations of E for the narrower and the wider sub-band, respectively. It is reassuring that the MW values based on all frames agree with those reported for the single frame considered in Section 4.1. More interestingly, there are two distinct regimes in the truncation error profiles. For both sub-bands, the largest error corresponding to the maximum peak of the mean curve is obtained by retaining only the zeroth-order BLCC. The means then become smaller and the spreads narrower as they turn into a flat regime of zero values. The turning point occurs near MW, beyond which the BLCC series may be assumed to contribute very little to the representation of the sub-band region.

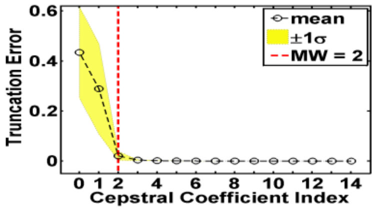

Figure 4: *Truncation error for the* C_l'*: [0.1,0.9]-kHz extracted at all frames of "IOWA".*

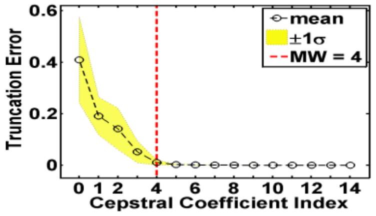

Figure 5: *Truncation error for the* C_l'*: [2.5,4.0]-kHz extracted at all frames of "IOWA".*

5. Summary and application possibilities

We have uncovered a linear relationship, which permits direct transformation of the Fourier series of full-band cepstral coefficients into an analogous series of band-limited cepstral coefficients (BLCCs). These represent the spectral region of the full band delimited by the selected sub-band. The parametric expression of the linear transformation gives the flexibility of selecting the left and right limits for any sub-band within the frequency range of the full band.

We have also provided some empirical justification for truncating the BLCC series after MW without significant loss in spectral resolution. This is therefore the upper bound N proposed for practical use, although a value just larger than MW might conceivably be necessary for a more exact representation of the sub-band spectrum.

The flexibility and efficiency afforded by our band-limiting method suggest that the resulting BLCCs have the potential to throw new light on some challenging problems in speech science and technology. For instance, BLCCs could facilitate further systematic studies of the sub-band dependence of speaker [4,5,6] and accent [7] variability. They could also play a major part in the quest for robust classification performance. There is increasing evidence that the use of sub-bands is indeed beneficial for automatic speech [8] and speaker classification [9,10,11,12] and, more recently, for simulated [13,14] and real-world [15] forensic voice comparison.

Finally, it has not escaped our notice that the Euclidean distance between two sets of BLCCs is a corollary of the mean square formula in Eq. (17). For $N = MW$, we conjecture that the BLCC-based distance measure will approach the "exact" measure proposed in [16], which implicates the entire set of M full-band coefficients regardless of the sub-band's width. This conjecture points to the possibility of improving the efficiency and perhaps even the statistical stability of band-limited cepstral distances in speech and speaker classifiers.

Table of Contents for this manuscript

6. Acknowledgements

I am grateful to Prof. Shunichi Ishihara for the uplifting discussions held locally in Canberra and remotely during his 2021-2022 sabbatical stay in Japan. I thank the anonymous reviewers for their suggestions. I extend sincere thanks to Prof. Phil Rose and Dr Yuko Kinoshita for their interest and queries.

7. References

[1] Deller Jr. J.R., Proakis, J.G. and Hansen, J.H.L., Discrete-Time Processing of Speech Signals, Macmillan, 1993.

[2] Picone, J.W., "Signal modelling techniques in speech recognition", *Proceedings of the IEEE*, 81(9): 1215-1247, 1993.

[3] Gradshteyn, I.S. and Ryshik, I.M., Tables of Integrals, Series and Products, Academic Press, 2007.

[4] Kitamura, T. and Akagi, M., "Speaker individualities in speech spectral envelopes", *Proc. 3rd International Conference on Spoken Language processing*, Yokohama, 1183-1186, 1994.

[5] Khodai-Joopari, M., Clermont, F. and Barlow, M., "Speaker variability on a continuum of spectral sub-bands from 297-speakers' non-contemporaneous cepstra of Japanese vowels", *Proc. 10th Australian International Conference on Speech Science and Technology*, Sydney, 504-509, 2004.

[6] Clermont, F., Kinoshita, Y. and Osanai, T., "Sub-band cepstral variability within and between speakers under microphone and mobile conditions: A preliminary investigation", *Proc. 16th Australasian International Conference on Speech Science and Technology*, Parametta, 317-320, 2016.

[7] Arslan, L.M. and Hansen, J.H.L, "A study of temporal features and frequency characteristics in American English foreign accent", *The Journal of the Acoustical Society of America,* 102(1): 28-40, 1997.

[8] Mokhtari, P. and Clermont, F., "Contributions of selected spectral regions to vowel classification accuracy", *Proc. 3rd International Conference on Spoken Language processing*, Yokohama, 1923-1926, 1994.

[9] Hayakawa, S. and Itakura, F., "Text-dependent speaker recognition using the information in the higher frequency band", *Proc. International Conference on Acoustics, Speech and Signal Processing,* Adelaide, 137-140, 1994.

[10] Finan, R.A, Damper, R.I. and Sapeluk, A.T., "Improved data modeling for text-dependent speaker recognition using sub-band processing", *International Journal of Speech Technology*, 4: 45-62, 2001.

[11] Sivakumaran, P., Ariyaeeinia, A.M. and Loomes, M.J., "Sub-band based text-dependent speaker verification", *Speech Communication*, 41: 485-509, 2003.

[12] Osanai, T., Kinoshita, Y. and Clermont, F., "Exploring sub-band cepstral distances for more robust speaker classification", *Proc. 17th Australasian International Conference on Speech Science and Technology*, Sydney, 41-44, 2018.

[13] Kinoshita, Y., Osanai, T. and Clermont, F., "Forensic voice comparison using sub-band cepstral distances as features: A first attempt with vowels from 306 Japanese speakers under channel mismatch conditions", *Proc. 17th Australasian International Conference on Speech Science and Technology*, Sydney, 45-48, 2018.

[14] Kinoshita, Y., Osanai, T., and Clermont, F., "Sub-band cepstral distance as an alternative to formants: Quantitative evidence from a forensic comparison experiment", 94, *Journal of Phonetics*, 2022.

[15] Rose, P., "Likelihood ratio-based forensic semi-automatic speaker identification with alveolar fricative spectra in a real-world case", *Proc. 18th Australasian International Conference on Speech Science and Technology*, Canberra, 2022.

[16] Clermont, F. and Mokhtari, P., "Frequency-band specification in cepstral distance computation", *Proc. 5th Australian International Conference on Speech Science and Technology,* Perth, I: 354-359, 1994.