Locating Spectral Regions of Speaker Sensitivity with the Sub-Band Mel-Frequency Cepstrum: An Investigation Guided by Forensic Voice Comparison

Frantz Clermont, Shunichi Ishihara

Speech and Language Laboratory, Australian National University, Canberra, Australia

dr.fclermont@gmail.com, shunichi.ishihara@anu.edu.au

Abstract

The non-uniform spectral encoding of speaker information suggests that certain sub-band regions carry strong sensitivity to that information. Such regions are sought here by: capturing locally-encoded spectral information with band-limited cepstral coefficients (BLCCs), and quantifying the sensitivity through forensic voice comparison (FVC). Using 5 Japanese vowels from 306 native male speakers, FVC experiments were conducted with BLCCs representing 18 sub-bands spaced across the full range [0-4000 Hz] on the mel scale. The FVC results highlight 3 speaker-sensitive sub-bands, located near formant regions in low-, mid-, and high-frequency ranges. Speaker information is strongly encoded in high-frequency sub-bands roughly above 2300 Hz.

Index Terms: Band-limited cepstral coefficients (BLCCs), mel-frequency scale, filter-bank, vowel, speaker sensitivity.

1. Introduction

Ever since the influential paper [1] which reported the superior performance of mel-frequency cepstral coefficients (MFCCs) in speech recognition, it has been argued that these acoustic parameters based on a logarithmic scale possess the significant advantage of increasing resolution in the lower frequency bands, while allowing "better suppression of insignificant spectral variation in the higher frequency bands" [1: 364]. MFCC spectra are thus enhanced in the lower bands that encode the bulk of phonetic differences. Yet, MFCC spectra have also been shown to perform well in speaker recognition [2,3].

The following question naturally arises: Which regions of MFCC spectra encode the bulk of speaker differences? We investigate this question with: (1) sub-band MFCCs to focus on local regions, one at a time, across the full band; and (2) forensic voice comparison (FVC) to quantify speaker sensitivity in each selected sub-band. This approach is put forward as a key to finding speaker-sensitive sub-bands, and potentially assisting forensic speech experts for a deeper analysis of FVC outcomes and a more logical validation of the adopted FVC system.

Sec. 2 describes the method [4] recently developed for obtaining sub-band MFCCs, also referred to as band-limited CCs (BLCCs) in this work. By contrast with the alternative method of repeating the spectrum-to-cepstrum conversion for every sub-band, we have adopted the BLCC method which can easily transform full-band into sub-band MFCCs with flexible sub-band selection. The sub-band results presented in Sec. 4 may thus be seen as supportive evidence for the BLCC method.

Sec. 3 outlines our FVC experiments and multi-speaker vowel data. Using 18 sub-bands spaced across the full band [0-4000 Hz] on the mel scale, the experiments were aimed at locating the vowel spectral regions most sensitive to speaker differences with FVC performance as a quantitative measure.

Sec. 4 gives a detailed map of the least and most speaker-sensitive sub-bands. The top 3 optimal sub-bands and the full band are also compared in terms of FVC performance. Sec. 5 discusses the main findings with suggestions for further study.

2. The BLCC method

Sec. 2.1 gives an overview of the method. Sec. 2.2 outlines the mathematical formulae, and Sec. 2.3 discusses the practical size for a BLCC vector.

2.1. Procedural steps

The BLCC method proceeds in 3 main steps shown in Fig. 1. Steps (1) and (2) describe standard procedures of spectral analysis applied to short-time frames of the speech signal with a sampling frequency F_s (Hz). The final step (3) consists of a linear transformation from full-band to sub-band cepstrum.

At Step (1), the log magnitude spectrum (LMS), $S(\omega)$, is computed from the energy outputs of a filter bank. The filters' centre frequencies are non-uniformly spaced across the full range $[0, (F_s/2)]$ in Hz (or $[0, \pi]$ in radians). The spacing follows the mel scale, roughly linear at low frequencies and nonlinear at high frequencies.

At Step (2), the Discrete Cosine Transform (DCT) is used to decorrelate $S(\omega)$ and reduce its dimensionality. The result is a Fourier series of cosine functions weighted by the so-called cepstral coefficients C_k. In this work, the C_k are referred to as full-band, mel-frequency CCs (MFCCs in short). The average of the full-band LMS is usually assumed to be zero, hence $C_0 = 0$. In practice, the series is truncated after $M \ll \infty$ terms:

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

Independently of the spacing on the frequency axis, the retention of a small number of terms carries a smoothing effect of its own, one which makes $S(\omega)$ more immune to "noninformation bearing variabilities" [5: 169].

At Step (3), BLCCs are obtained using a new method which gives the flexibility of selecting a sub-band region of the full-band spectrum *without having to repeat the previous steps.* As shown in [4], the vector $\mathbf{c}'$ of BLCCs representing a sub-band can indeed be calculated *directly* from the vector $\mathbf{c}$ of full-band C_k using a linear transformation $\mathbf{A}$ expressed in Sec. 2.2.

2.2. Linear transformation formulae

A sub-band region $[\omega_1, \omega_2]$ of the cepstrally-smoothed, full-band LMS is represented by the Fourier cosine series given in Eq. (2). The mathematical aim is to express the band-limited coefficients C'_l as functions of the full-band C_k.

$$S(\omega(\omega')) \cong C'_0 + \sum_{l=1}^{N} C'_l \cos(l\omega'), \; 0 \le \omega' \le \pi, \tag{2}$$

where C'_l is the l-th BLCC and N is the series' upper bound.

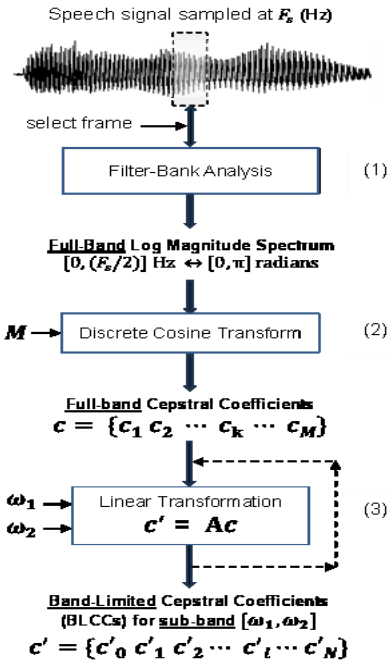

Figure 1: *The BLCC method and its main steps*.

Note that Eq. (2) includes C'_0 because the average of $S(\omega(\omega'))$ within a sub-band may not be zero. The other $C'_{l>0}$ account for the spectral shape within the selected sub-band.

The frequency variable ω' defined below translates the sub-band interval $[\omega_1, \omega_2]$ to that of the full-band range $[0, \pi]$:

$$\omega' = \pi \left[\frac{(\omega-\omega_1)}{(\omega_2-\omega_1)}\right],\ \omega_1 \le \omega \le \omega_2 \tag{3}$$

From Eq. (3) the frequency variable ω of the full-band series can be expressed as follows, where the scalar W is the ratio of the sub-band's width to the full-band's frequency range:

$$\omega(\omega') = \omega_1 + \left[\frac{(\omega_2-\omega_1)}{\pi}\right]\omega' = \omega_1 + W\omega' \tag{4}$$

The notation $\omega(\omega')$ is a reminder that ω is a (band-dependent) function of ω'. It is thus possible to substitute ω in Eq. (1) for Eq. (4), and use standard formulae to obtain a *linear* relation between the band-limited C'_l (Eq. 2) and full-band C_k:

$$C'_l = \sum_{k=1}^{M} a_{lk}\, C_k,\ l = 0,1,\dots,N \tag{5}$$

Eq. (5) leads to the matrix form $\mathbf{c}' = \mathbf{A}\mathbf{c}$, where $\mathbf{A}$ is the transformation matrix for the selected sub-band, $\mathbf{c}'$ the column vector of C'_l, and $\mathbf{c}$ the column vector of C_k. The elements of $\mathbf{A}$ are defined in Eq. (6a) for $l = 0$ and Eqs. (6b)-(6c) for $l > 0$:

$$a_{lk,\, l=0} = \beta_k[\sin(k\omega_2) - \sin(k\omega_1)] \tag{6a}$$

$$a_{lk,\, l\neq kW} = \gamma_{lk}[(-1)^{l+1}\sin(k\omega_2) + \sin(k\omega_1)] \tag{6b}$$

$$a_{lk,\, l=kW} = \cos(k\omega_1) \tag{6c}$$

where:

$$\beta_k = \frac{1}{k(\omega_2-\omega_1)} \text{ and } \gamma_{lk} = \frac{2(kW)}{\pi[l^2-(kW)^2]} \tag{6d}$$

2.3. BLCC vector: Practical size and spectral resolution

The Fourier-series (BLCC) representation of a sub-band region will theoretically improve with increasing values of the upper bound N. How large does N need to be in practice? The solution proposed in [4] is to truncate the BLCC series after $N = M \times W$ (MW in short) terms: M is the size of the full-band vector of C_k, and W (defined earlier) is the fraction of the full-band's range occupied by the sub-band's width. Note that MW will generally not be an integer; it must therefore be rounded for practical use.

Our rationale for MW is this: If M cepstral coefficients represent the full-band spectrum with a certain resolution, then roughly the same resolution for a sub-band region should be achievable with $C'_{l=0,1,\dots N=MW}$. That is, for N fixed at around MW, there should be no significant loss in spectral resolution. The numerical profiles of BLCCs illustrated in [4,6] show that the dominant coefficients indeed extend up to MW, followed by a noticeable decay of their magnitude towards zero.

3. Experimental procedure

3.1. Speech materials and parametrisation

The speech materials were sourced from the Japan's National Research Institute of Police Science (NRIPS) database [7], consisting of microphone recordings from 306 adult-male, native Japanese speakers (mean age: 39.9, SD: 15.5). The 5 Japanese vowels (/a, e, i, o, u/ in null consonantal context) selected for this work were uttered twice in 2 sessions split 3-5 months apart. A sampling rate of 8000 Hz was used to contain the full frequency range within the Japanese telephone bandwidth [0-4000 Hz]. This bandwidth constraint combined with the non-contemporaneous recordings satisfy some of the basic requirements for forensic relevance.

Using 25 triangular-shaped filters spaced evenly on the mel scale with 50% overlap, filter-bank analysis was applied to each vowel segment using a frame size of 25 msecs and a step size of 5 msecs. For each frame, 14 MFCCs were obtained by performing DCT on the filter's log energy outputs. The MFCCs averaged over the 3 middle frames were retained as the vector of full-band MFCCs.

Subsequently, the full frequency range [0-4000 Hz] was scanned with a 600-Hz sub-band, shifted every 200 Hz, and the vector of BLCCs was obtained for each sub-band by linear transformation of the full-band MFCCs as described in Sec. 2. There are 18 sub-bands in total, resulting in 18 corresponding vectors of BLCCs. Applying the formula in Sec. 2.3, MW equals 2.1, rounded to 2. The size of each BLCC vector is therefore 3, including the 0th-order coefficient.

3.2. Data partitioning and experimental descriptions

The speech materials were randomly divided into 3 batches, which were rotated and used as the test, background, and calibration databases, resulting in 6-fold cross-validation FVC experiments (see Table 1). The results presented in this paper are the averages of these 6 experiments. There are 612 same-speaker comparisons and 61,812 different-speaker comparisons from these cross-validation experiments.

Table 1: *Six-fold cross-validation experiments*

	Test	Background	Calibration
1	Batch 1	Batch 2	Batch 3
2	Batch 1	Batch 3	Batch 2
3	Batch 2	Batch 3	Batch 1
4	Batch 2	Batch 1	Batch 3
5	Batch 3	Batch 2	Batch 1
6	Batch 3	Batch 1	Batch 2

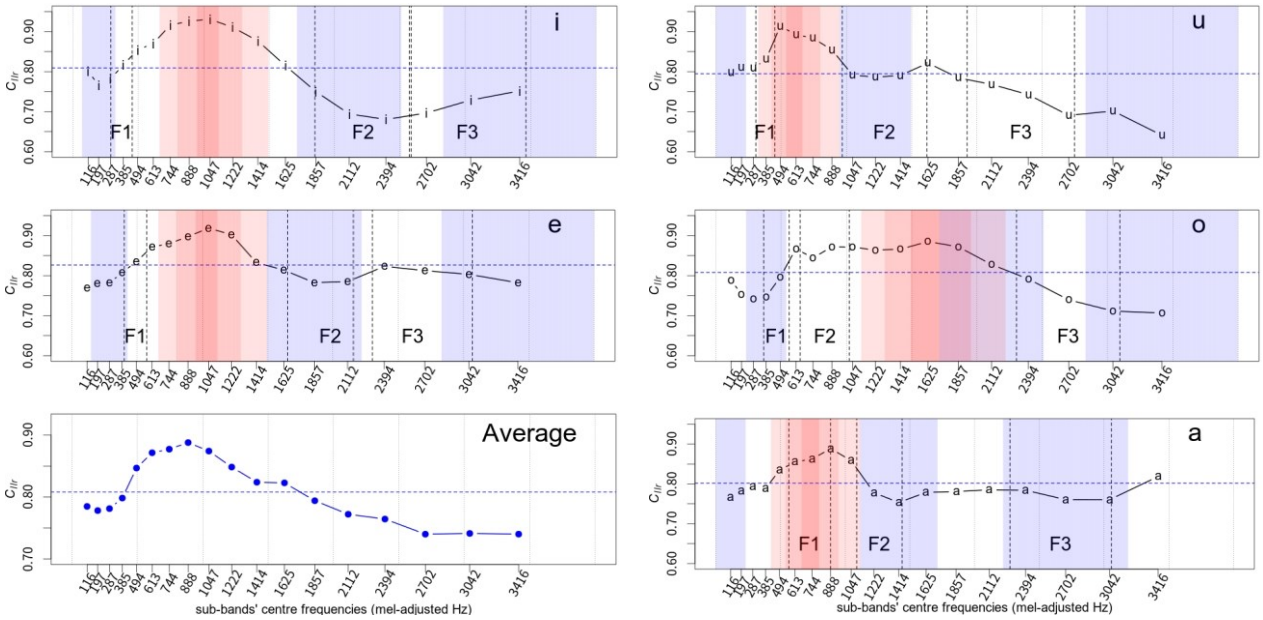

Figure 2: *Results for each of the 5 Japanese vowels and for all vowels are displayed in separate panels: C_{llr} values (y-axis) for the 18 sub-bands (x-axis) are connected by a black curve, and the vowel-averaged C_{llr} values by a blue curve. In every panel, the horizontal dashed line (in blue) indicates the C_{llr} value averaged over the 18 sub-bands. The 3 best-performing sub-bands (fused) are filled in blue, while the 3 under-performing sub-bands (fused) are filled in pink. In each vowel panel, the vertical dotted lines mark F1, F2 and F3 regions ±1 SD away from the speaker-averaged means.*

3.3. Experimental descriptions

Two FVC experiments were conducted on a per-vowel basis, with BLCCs representing each of the 18 sub-bands. Likelihood Ratios (LRs) were obtained using the Multivariate Kernel Density model [8], followed by logistic regression calibration [9]. The performance of the FVC system was assessed using the log-LR cost (C_{llr}) for each sub-band. *The closer the C_{llr} is to 0 and below 1, the better the performance of the FVC system*.

Experiment 1 sought to uncover how speaker information is encoded across different spectral regions for each vowel. If a sub-band is more sensitive to speaker information, the FVC system performs better with its BLCCs than with those from other sub-bands. Experiment 2 systematically combined sub-band LRs from Experiment 1, and compared the performance with full-band MFCCs results to locate the optimal sub-bands.

3.4. Sub-band limits on mel-frequency scale

Although the 18 sub-bands were selected on an equally-spaced Hz scale to maintain the uniform analytical potential of each sub-band, their corresponding C_{llr} values need to be positioned with respect to the non-linear scale on which the full-band MFCCs are based. To this end, the frequency axis of the C_{llr} graphs was mapped onto a mel-adjusted Hz scale. The adjusted limits $[\omega_1, \omega_2]$ for the 18 sub-bands are given in Table 2.

Table 2: *Sub-bands' mel-adjusted limits (Hz).*

1	[0, 231]	2	[70, 324]	3	[147, 427]
4	[231, 539]	5	[324, 663]	6	[427, 799]
7	[539, 949]	8	[663, 1114]	9	[799, 1295]
10	[949, 1494]	11	[1114, 1714]	12	[1295, 1955]
13	[1494, 2220]	14	[1714, 2511]	15	[1955, 2832]
16	[2220, 3185]	17	[2511, 3573]	18	[2832, 4000]

This adjustment is also desirable because it facilitates a more intuitive interpretation of sub-band locations, and a direct comparison with analogous results from previous studies.

4. Results and discussion

The results of Experiments 1 and 2 are depicted jointly in Fig. 2, but discussed separately in Secs. 4.1 and 4.2, respectively. The F1, F2, and F3 regions (means ±1 SD) for each vowel are marked by vertical dotted lines. Formant information is provided for reference only. See [10] for details on the formant measurement procedure.

The C_{llr} values (y-axis) from Experiment 1 are plotted in Fig. 2 against the sub-bands' centre frequencies (x-axis) for each vowel (see panels labelled "i", "u", "e", "o", and "a"). The C_{llr} values averaged over all vowels are shown in the panel labelled "Average". The average C_{llr} value of the 18 data points in each panel is shown as a dashed horizontal line to help the visual comparison of C_{llr} fluctuations from vowel to vowel.

Note that the centre frequencies are the midpoints of the sub-band limits given in Table 2. Following the non-linear property of the mel scale, the spacing between the 18 data points plotted in Fig. 2 becomes wider as the frequency increases towards the higher spectral regions.

4.1. Experiment 1

All C_{llr} values in Fig. 2 remain consistently below 1, suggesting that every sub-band carries some useful speaker information. The fluctuations in C_{llr} values clearly indicate that speaker information is non-uniformly distributed across the full band, a phenomenon observed in previous studies [11-15]. Owing to its flexible sub-band selection and computational efficiency, our BLCC method has facilitated the task of creating the detailed map of speaker-sensitive sub-bands superposed in Fig. 2.

The C_{llr} patterns in the vowel panels of Fig. 2 are consistent as far as relating local extrema to sub-band regions that are more or less sensitive to speaker information: (1) the relatively high C_{llr} values (local maxima) point to the least speaker-sensitive sub-bands (*filled in pink*) ranging roughly between 300 Hz and 2200 Hz; (2) the relatively low C_{llr} values (local minima) indicate the most speaker-sensitive sub-bands (*filled in blue*) located around low-, mid- and high-frequency ranges.

Overall, speaker information appears to be more strongly encoded towards the *high-frequency* sub-bands, as depicted in the Average panel where the C_{llr} pattern drops to its minimum from about 2300 Hz. While the *low-frequency range* (roughly below 500 Hz) tends to be less sensitive than the high-frequency one, it does carry notable speaker information albeit with vowel-to-vowel variations in C_{llr} values. Previous studies [6,13] have highlighted the relevance of this low range for speaker classification, especially targeting the back vowels of Japanese. In a similar vein, [16] have reported that the range [100-300 Hz] carries significant speaker information based on Japanese sentences spoken at normal, slow and fast rate, and [17] have found the range [0-770 Hz] to be highly useful for idendifying speakers with various accents of British English.

Fig. 2 also shows how the speaker-sensitive sub-bands identified through FVC are positioned with respect to formant regions. While exact alignment may not be expected partly due to statistical uncertainty in formant measurement, the speaker-sensitive sub-bands tend to fall near or within the marked F1, F2 and F3 regions and, in most cases, extend into likely F4 regions. This trend is reassuring with regard to the formant-dependent properties of vowels but, more importantly, it underscores the advantage of using the BLCC method to focus on any sub-band regions of potential relevance to speaker-specific information.

4.2. Experiment 2

In addition to identifying the locations of speaker-sensitive sub-bands as per the above, it is of interest to gauge their relative importance and performance with respect to the full band.

Experiment 2 thus consisted of fusing LRs for combinations of 2 to 7 sub-bands, with a view to observing FVC performance with an increasing number of sub-bands, and determining the combinations that yield optimal results. The best C_{llr} values per vowel are plotted in Fig. 3 for single sub-bands at left and, then, for fused sub-bands. The major improvements in C_{llr} occur with 3 fused sub-bands corresponding to those highlighted (in blue) in Fig. 2, in the same order from low to high-frequency regions.

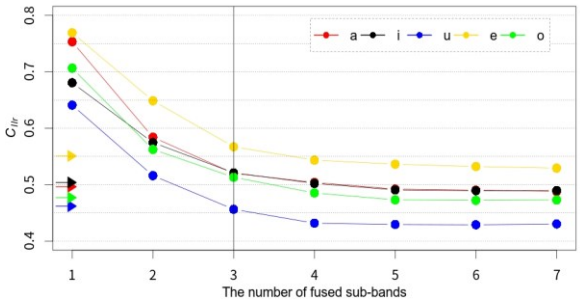

Figure 3: *Per-vowel Cllr values for the most speaker-sensitive sub-bands: single (1) and fused (2-7). Left arrows indicate Cllr values at full-band.*

The vertical line in Fig. 3 marks the point at which the inclusion of high-frequency sub-bands causes C_{llr} values to approach those obtained at full band, thus reinforcing our earlier observation regarding the significant role of those sub-bands. The speaker sensitivity of the high-frequency regions of our cepstrally-smoothed spectra agrees with previous acoustical studies of Japanese [13,18] and English [15,19] vowels. Shape differences in the lower part of the vocal tract (laryngeal tube and piriform fossa) are thought to cause the high-frequency range of speaker variation [20].

5. Concluding discussion and future work

This work has illustrated the efficacy of mel-frequency BLCCs and the usefulness of FVC in investigating sub-band regions of vowel spectra that are sensitive to speaker diffrences.

There is a clear consistency in the overall distributional pattern of speaker sensitivity and the locations of the most informative sub-bands (fused) with respect to relatively low C_{llr} values. For all vowels, two of these sub-bands respectively span the low- and the high-frequency ranges, while the other sub-band tends to be around the middle of the full range.

The relative performance of these sub-bands (quantified in experiment 2) sheds some light on the question raised in the introduction regarding MFCC spectra. Speaker information is clearly detectable in low- and mid-frequency ranges where spectral resolution is enhanced on the mel scale. More notably, our results indicate the effectiveness of MFCCs in capturing strong speaker differences in the vowels' high-frequency regions. Yet, this is where the frequency resolution of the mel scale is lower, and where spectral distinctiveness between speakers would expectedly be reduced. This apparent contradiction warrants further study, especially since the success of MFCCs in speech and speaker classification is commonly attributed to the very properties of the mel scale.

To that end, it may be beneficial to repeat the same FVC experiments using BLCCs on alternative frequency scales that differ in filter allocation: (1) BLCCs extracted from inverse-MFCCs [21], i.e., with the filters at high frequencies shifted to low frequencies, and vice versa; and (2) BLCCs extracted from linear-frequency CCs (LFCCs), i.e., with uniformly-spaced filters. In either case, it may be worthwhile to increase the number of filters (set to 25 in this work) and observe the impact of enhanced frequency resolution throughout the full range. Ultimately, some valuable insights may be gained from the patterns of speaker variances (within and between) in high-frequency sub-bands spaced on linear and mel scales.

The current study can conceivably be extended to non-vowel sounds [22] and female speech [23], which have been less studied in forensic context. The BLCC method should also be useful for locating the sub-bands that are more robust under adverse conditions, such as with mobile phones [24] and noisy environments, or for studying the impact of emotion-dependent sub-bands [25] on speaker variability.

Finally, it is hoped that the sub-band approach embedded in BLCCs will provide new perspectives in other areas of speech science and technology, where detailed investigation of spectral regions is required in an efficient and flexible manner. Possible areas of application extend from acoustic phonetics (e.g., the investigation of contrastive features [26,27]) and socio-phonetics (e.g., exploration of accent-specific sub-bands [28]) to spoofing detection [29,30], where synthetic traits may be notably encoded in different spectral regions.

6. Acknowledgements

The authors thank two anonymous reviewers for their valuable comments.

7. References

[1] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357-366. 1980.

[2] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition from features to supervectors," *Speech Communication*, vol. 52, pp. 12-40. 2010.

[3] S. S. Tirumala, S. R. Shahamiri, A. S. Garhwal and R. Wang, "Speaker identification features extraction methods: A systematic review," *Expert Systems with Applications*, vol. 90, pp. 250–271. 2017.

[4] F. Clermont, "Linear transformation from full-band to sub-band cepstrum," in, *Proceedings of the 18th Australasian International Conference on Speech Science and Technology*, 2022, pp. 136-140.

[5] L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, 1st ed. (Prentice Hall signal processing series). Englewood Cliffs, N.J.: Prentice Hall, 1993.

[6] S. Ishihara and F. Clermont, "The sub-band cepstrum as a tool for local spectral analysis in forensic voice comparison," in, *Proceedings of the 21st Annual Workshop of the Australasian Language Technology Association*, 2023, pp. 40-50.

[7] H. Makinae, T. Osanai, T. Kamada, and M. Tanimoto, "Construction and preliminary analysis of a large-scale bone-conducted speech database," *The Institute of Electronics, Information and Communication Engineers (IEICE) Technical Report*, vol. 107, no. 165, pp. 97-102, 2007.

[8] C. G. G. Aitken and D. Lucy, "Evaluation of trace evidence in the form of multivariate data," *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, vol. 53, no. 1, pp. 109-122. 2004.

[9] G. S. Morrison, "Tutorial on logistic-regression calibration and fusion: Converting a score to a likelihood ratio," *Australian Journal of Forensic Sciences*, vol. 45, no. 2, pp. 173-197. 2013.

[10] Y. Kinoshita, T. Osanai, and F. Clermont, "Sub-band cepstral distance as an alternative to formants: Quantitative evidence from a forensic comparison experiment," *Journal of Phonetics*, vol. 94, no. 101177. 2022.

[11] G. E. Peterson, "The acoustics of speech–part II: Acoustical properties of speech waves." In L. E. Travis (ed.), *Handbook of Speech Pathology*, 1st ed., pp. 137-173. New York: Appleton-Century-Crofts, Inc., 1995.

[12] T. Kitamura and M. Akagi, "Speaker individualities in speech spectral envelopes," *Journal of the Acoustical Society of Japan (E)*, vol. 16, no. 5, pp. 283-289. 1995.

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

[14] R. Goto, K. Misawa and Y. Okada, "Analysis of individual characteristics in vowel spectral envelopes," In, *Proceedings of the International Multi-Conference of Engineers and Computer Scientists*, 2017, pp. 113-116.

[15] P. Mokhtari and F. Clermont, "A methodology for investigating vowel-speaker interactions in the acoustic-phonetic domain," in, *Proceedings of the 6th Australian International Conference on Speech Science and Technology*, 1996, pp. 127-132.

[16] X. Lu and J. Dang, "An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification," *Speech Communication*, vol. 50, pp. 312–322. 2008.

[17] S. Safavi, A. Hanani, M. Russell, P. Jančovič and M. Carey, "Contrasting the effects of different frequency bands on speaker and accent identification," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 829-832. 2012.

[18] T. Kitamura and M. Akagi, "Relationship between physical characteristics and speaker individualities in speech spectral envelopes", in *Proceedings of 3rd Joint Meeting of the Acoustical Society of Japan and the Acoustical Society of Japan*, 1996, pp. 833-837.

[19] M. Sambur, "Selection of acoustic features for speaker identification," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 23, no. 2, pp. 176-182. 1975.

[20] T. Kitamura, K. Honda and H. Takemoto, "Individual variation of the hypopharyngeal cavities and its acoustic effects", in *Acoustical Science & Technology*, 2005, pp. 16-25.

[21] H. Lei and E. Lopez-Gonzalo, "Mel, linear, and antimel frequency cepstral coefficients in broad phonetic regions for telephone speaker recognition," in, *Proceedings of Inerspeech 2009*, 2009, pp. 2323-2326.

[22] P. Rose, "Likelihood ratio-based forensic semi-automatic speaker identification with alveolar fricative spectra in a real-world case," in, *Proceedings of the 18th Australasian International Conference on Speech Science and Technology*, 2022, pp. 6-10.

[23] P. Rose and E. Winter, "Traditional forensic voice comparison with female formants: Gaussian mixture model and multivariate likelihood ratio analyses," in, *Proceedings of the 13th Australasian International Conference on Speech Science and Technology*, 2010, pp. 42-15.

[24] B. B. T. Nair, E. A. S. Alzqhoul and B. J. Guillemin, "Impact of the GSM and CDMA mobile phone networks on the strength of speech evidence in forensic voice comparison," *Journal of Forensic Research*, vol. 7, no. 324, pp. 1-9. 2016.

[25] M. H. Abed and D. Sztahó, "Effects of emotional speech on forensic voice comparison using deep speaker embeddings," in, *Proceedings of the XIX Hungarian Computational Linguistics Conference*, 2023, pp. 159-170.

[26] K. Iskarous, "The encoding of vowel features in mel-frequency cepstral coefficients." *Il parlato nel contesto naturale* [Speech in the Natural Context], 2018, pp. 9-18.

[27] M. Lambropoulos, F. Clermont and S. Ishihara, "The sub-band cepstrum as a tool for locating local spectral regions of phonetic sensitivity: A first attempt with multi-speaker vowel data," in, *Proceedings of Interspeech 2024*, 2024, pp. 1535-1539.

[28] L. M. Arslan and J. H. L. Hansen, "A study of temporal features and frequency characteristics in American English foreign accent," *The Journal of the Acoustical Society of America*, vol. 102, no. 1, pp. 28-40. 1997.

[29] M. H. Soni, T. B. Patel and H. A. Patil, "Novel subband autoencoder features for detection of spoofed speech," in, *Proceedings of Interspeech 2016*, 2016, pp. 1820-1824.

[30] B. Chettri, T. Kinnunen and E. Benetos, "Subband modeling for spoofing detection in automatic speaker verification," in, *Proceedings of Odyssey 2020: The Speaker and Language Recognition Workshop*, 2020, pp. 341-348.