Letter to the Editor: *Critique* of the Article *Linear Transformation from Full-Band to Sub-Band Cepstrum* by F. Clermont in Proceedings SST2022

Michael Wagner

NCBS Pty Ltd, University of Canberra, Technical University Berlin

michael.wagner@ncbs.com.au

Abstract

This letter to the Editor discusses the paper *Linear Transformation from Full-Band to Sub-Band Cepstrum* by F. Clermont, published in the *Proceedings of SST2022*. It argues that the concept of *sub-band cepstrum* is not defined in the paper or in the references. It also argues that the method proposed in the paper is essentially band-pass filtering the cepstrum and as such is ineffective because it only sets stop band log-magnitude spectral values to zero, which does not eliminate or even necessarily attenuate the stop band. An improved method which uses logarithmic attenuation of the stop band is suggested instead. Further, this letter argues that the operational equation effecting the cepstral band-pass filtering is flawed and offers an alternative.

Index Terms: sub-band cepstrum, band-pass filter, speech signal processing.

1. Introduction

In an article entitled *Linear Transformation from Full-Band to Sub-Band Cepstrum* by F. Clermont, published in the Proceedings of SST2022 [1], the author proposes a method to obtain a *sub-band cepstrum*. Several problematic aspects of that proposal are addressed by this letter, including the failure to define, or provide a reference to, the concept of *sub-band cepstrum*; the impossibility of determining a band-pass cepstrum in the strict sense of eliminating the influence of the stop band; and a significant error in the mathematical justification of the method proposed in the article.

2. The *Critique*

2.1. Missing definition and BP filter assumption

The article does not provide a definition of *sub-band cepstrum*; instead, the first sentence states "Sub-band cepstra are commonly estimated using the filterbank method [Ref. 1, Ref. 2]"[1], but neither Ref. 1 nor Ref. 2 contains a definition or even a mention of *sub-band cepstrum*.

However, both Ref. 1 and Ref. 2 contain sections with detailed discussions of band-pass filtering of time-domain signals, which eliminates from the signal the influence of a part of the spectrum that is defined as the *stop band*. It is therefore plausible that the unstated intention of the article is to apply the equivalent of a band-pass filtering method that is well established for time-domain signals to effectively band-pass filter a cepstrum.

Therefore, this letter now proceeds to examine the method proposed in the article on the assumption that the purpose of that method is to eliminate from the cepstrum the influence of a spectral stop band.

[1] Ref. 1 and Ref. 2 are numbered [2] and [3], respectively, in the list of References below.

2.2. Is band-pass filtering of a cepstrum possible?

In digital signal processing, band-pass filtering of a signal $x(n)$ is usually a 3-step process:

1. The signal is transformed into the spectral domain by an operator T: $x(n) \rightarrow X(k)$.
2. The signal spectrum $X(k)$ is multiplied with a rectangular window $H(k)$ to become $Y(k) = H(k) \cdot X(k)$.
3. The pass-band spectrum $Y(k)$ is transformed back into the signal domain by the inverse operator T^{-1}: $Y(k) \rightarrow y(n)$.

The operator T can be, for example, a discrete Fourier transform, a discrete cosine transform or, as proposed in the article, a Fourier cosine series, all of which have inverse operators T^{-1}. An example for $H(k)$ is shown here:

$$H(k) = \begin{cases} 1, \text{ for } |k - k_0| \leq B/2 \\ 0, \text{ for } |k - k_0| > B/2 \end{cases} \quad (1)$$

for a pass band with centre frequency k_0 and bandwidth B.

If this analysis method is applied to a cepstrum $x(n)$ instead of a time-domain signal, several observations are immediately obvious: firstly, the signal spectrum $X(k) = T[x(n)]$ that corresponds to the cepstrum is actually a log-magnitude spectrum – not a magnitude spectrum – of the original time-domain signal; hence, secondly, the function $H(k)$ is linear only with respect to the log-magnitude spectrum of the time-domain signal, not with respect to the magnitude spectrum; and thirdly, the function $H(k)$ does not eliminate, or necessarily even attenuate, the spectral magnitudes $Y(k)$ in the stop band. This implies that band-pass filtering of a cepstrum can never create a zero-magnitude stop band. A band-pass filtered cepstrum in the strict sense of the phrase is an impossibility.

However, a reasonable and practicable alternative for the function $H(k)$ that will guarantee the attenuation of spectral log-magnitudes over the chosen stop band is the *logarithmic-attenuation function*

$$Y(k) \equiv \mathrm{F}[X(k)] = \begin{cases} X(k), & \text{for } |k - k_0| \leq B/2 \\ X(k) - D, & \text{for } |k - k_0| > B/2 \end{cases} \quad (2)$$

with, for example, a constant $D = 3$ for a 30 dB spectral log-magnitude reduction over the stop band.

2.3. The mathematical justification of the method

The mathematical justification of the proposed method begins with the representation of a continuous log-magnitude spectrum

$$S(\omega) = \sum_{n=1}^{M} x(n) \cos(n\omega), \quad 0 \leq \omega \leq \pi, \quad (3)$$

as a truncated Fourier cosine series of discrete cepstrum values $x(n)$, where the order M of the truncation is to yield a suitably smoothed log-magnitude spectrum.

The article does not discuss the necessary conditions for that representation, which include evenness $[S(\omega) = S(-\omega)]$ and periodicity $[S(\omega) = S(\omega + 2\pi)]$, but in this instance, those preconditions are plausibly met for an ordinary discrete cepstrum $x(n)$, so that Eq. 3 is a valid instance of the operator T in Step 1 of the generic band-pass filtering process in Subsection 2.2 of this letter. This confirms our assumption in Subsection 2.1 that the proposed method is intended to effectively band-pass filter a cepstrum.

With a variable substitution of $\omega' = (\omega - \omega_1)/(\omega_2 - \omega_1)$, it is then claimed in Subsection 3.2 of the article that a true subset $\{S(\omega), \omega_1 \leq \omega \leq \omega_2\} \subset \{S(\omega), 0 \leq \omega \leq \pi\}$ could be represented as a different truncated Fourier cosine series

$$S(\omega(\omega')) = \sum_{n=0}^{M} x'(n) \cos(n\omega'), 0 \leq \omega' \leq \pi, \quad (4)$$

Eq. 4 being called a *band-limited analogue* of Eq. 3. However, the analogy is false, because in spite of the variable substitution, the subset is neither even $[S(\omega_1 + \omega) \neq S(\omega_1 - \omega)]$ nor periodic $[S(\omega) \neq S(\omega + 2(\omega_2 - \omega_1))$, thereby invalidating Eq. 4 as the operator T in Step 1 of a band-pass filtering process.

Fig. 1 illustrates the fundamental problem of cepstral band-pass filtering by showing the filtering of the cepstrum of a 20 ms vowel frame near the first vowel target of the diphthong “o” in “hello”. Fig. 1a shows the log-magnitude spectrum $X(k)$ of Step 1 of the 3-step process and indicates a 300 Hz sub-band [1.250 kHz, 1.550 kHz] that may be of particular interest. Fig. 1b shows the windowed log-magnitude spectrum $Y(k) = H(k) \cdot X(k)$ of Step 2 of the process. The lower part of the stop band is amplified to a constant $\log|Y(k)| \equiv 0$, and the upper part is part-amplified and part-attenuated to the same value. Neither the lower nor the upper part of the stop-band *magnitude* spectrum is eliminated, and the inverse operator T^{-1} which integrates over the entire range of the windowed log-magnitude spectrum in Step 3, would clearly be dominated by the stop band rather than by the sub-band of interest.

Consequently, it is difficult to endorse a claim in the abstract of the article that the “paper demonstrates the possibility of estimating the cepstrum for a sub-band region of the full-band spectrum” or a claim in the Summary Section 5 of the article that “the band-limited cepstral coefficients […] represent the spectral region of the full band delimited by the selected sub-band.”

To rectify that process, we suggest that Eq. 3 should be used as the operator T in Step 1 of the 3-step process, the function $\mathrm{F}[X(k)]$ in Eq. 2 should be used in Step 2, and the “standard Fourier integration formulae”, described in Subsection 3.3.1. of the article, should be used for Step 3 of the process. This process would still not be a cepstral band-pass filter in the strict sense of eliminating the influence of the stop band, but it is guaranteed to attenuate the stop band to any desired degree by logarithmic attenuation.

3. Conclusions

This letter argues that the article does not provide a definition or a reference to a definition of the concept of *sub-band cepstrum*.

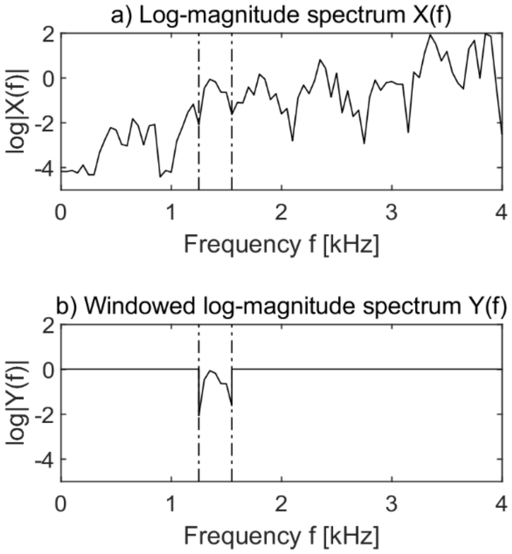

Figure 1. a) log-magnitude spectrum $X(f)$ and target sub-band; b) windowed log-magnitude spectrum $Y(f)$.

We have argued, however, that the cited references [2, 3], the initial Eqs 3 and 4 and the general descriptions of the method indicate that it is modelled on the well-established band-pass filtering of time-domain speech signals, except that it operates on cepstra instead of speech waveforms. Secondly, we have argued that band-pass filtering of a cepstrum and the term *sub-band cepstrum* may mislead the reader because unlike the band-pass filtering of a speech waveform, band-pass filtering of a cepstrum does not eliminate, or even necessarily attenuate, the influence of the stop band since it is the log-magnitude spectral values of the stop band that are set to zero and not the magnitudes. Strictly speaking, a band-pass-filtered cepstrum is an impossibility. Finally, we have shown that the central Eq. 4 in the article is incorrectly claimed by the article to be the equivalent of Eq. 3 applied to a log-spectral sub-band because, unlike Eq. 3, Eq. 4 does not meet the necessary preconditions of evenness and periodicity for a Fourier cosine series representation.

Two alternative methods have been suggested in this letter: firstly, to use a *logarithmic-attenuation function* instead of the effective rectangular band-pass window proposed in the article and, secondly, to replace the proposed method that is based on the flawed Eq. 4 of the article with the correct 3-step band-pass filtering process, best in conjunction with the *logarithmic-attenuation function*.

4. References

[1] Clermont, F., “Linear Transformation from Full-Band to Sub-Band Cepstrum”, Proc. Austral. Int. Conf. on Speech Science and Technology, SST2022, ASSTA Inc, 2022, pp 136-140.

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

[3] Picone, J.W., “Signal Modelling Techniques in Speech Recognition”, Proc. IEEE, 81(9): 1215-1247, 1993.