

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.718

Corrigendum 3

(01/2011)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA,
DIGITAL SYSTEMS AND NETWORKS

Digital terminal equipments – Coding of voice and audio
signals

Frame error robust narrow-band and wideband
embedded variable bit-rate coding of speech and
audio from 8-32 kbit/s

Corrigendum 3: Corrections to text and C-code

Recommendation ITU-T G.718 (2008) – Corrigendum 3

ITU-T G-SERIES RECOMMENDATIONS
TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
TRANSMISSION MEDIA AND OPTICAL SYSTEMS CHARACTERISTICS	G.600–G.699
DIGITAL TERMINAL EQUIPMENTS	G.700–G.799
General	G.700–G.709
Coding of voice and audio signals	G.710–G.729
Principal characteristics of primary multiplex equipment	G.730–G.739
Principal characteristics of second order multiplex equipment	G.740–G.749
Principal characteristics of higher order multiplex equipment	G.750–G.759
Principal characteristics of transcoder and digital multiplication equipment	G.760–G.769
Operations, administration and maintenance features of transmission equipment	G.770–G.779
Principal characteristics of multiplexing equipment for the synchronous digital hierarchy	G.780–G.789
Other terminal equipment	G.790–G.799
DIGITAL NETWORKS	G.800–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999
MULTIMEDIA QUALITY OF SERVICE AND PERFORMANCE – GENERIC AND USER-RELATED ASPECTS	G.1000–G.1999
TRANSMISSION MEDIA CHARACTERISTICS	G.6000–G.6999
DATA OVER TRANSPORT – GENERIC ASPECTS	G.7000–G.7999
PACKET OVER TRANSPORT ASPECTS	G.8000–G.8999
ACCESS NETWORKS	G.9000–G.9999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T G.718

Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s

Corrigendum 3

Corrections to text and C-code

Summary

Recommendation ITU-T G.718 describes a narrow-band (NB) and wideband (WB) embedded variable bit-rate coding algorithm for speech and audio operating in the range from 8 to 32 kbit/s which is designed to be robust to frame erasures.

This codec provides state-of-the-art NB speech quality over the lower bit rates and state-of-the-art WB speech quality over the complete range of bit rates. In addition, the ITU-T G.718 codec is designed to be highly robust to frame erasures, thereby enhancing the speech quality when used in IP transport applications on fixed, wireless and mobile networks. Despite its embedded nature, the codec also performs well with both NB and WB generic audio signals.

This codec has an embedded scalable structure, enabling maximum flexibility in the transport of voice packets through IP networks of today and in future media-aware networks. In addition, the embedded structure of ITU-T G.718 will easily allow the codec to be extended to provide a super-wideband and stereo capability through additional layers which are currently under development. The bitstream may be truncated at the decoder side or by any component of the communication system to instantaneously adjust the bit rate to the desired value without the need for out-of-band signalling. The encoder produces an embedded bitstream structured in five layers corresponding to the five available bit rates: 8, 12, 16, 24 and 32 kbit/s.

The ITU-T G.718 encoder can accept WB sampled signals at 16 kHz, or NB signals sampled at either 16 or 8 kHz. Similarly, the decoder output can be 16 kHz WB, in addition to 16 or 8 kHz NB. Input signals sampled at 16 kHz, but with bandwidth limited to NB, are detected by the encoder.

The output of the ITU-T G.718 codec is capable of operating with a bandwidth of 300-3400 Hz at 8 and 12 kbit/s and 50-7000 Hz from 8 to 32 kbit/s.

The high quality codec core represents a significant performance improvement, providing 8 kbit/s wideband clean speech quality equivalent to the ITU-T G.722.2 codec at 12.65 kbit/s whilst the 8 kbit/s narrow-band codec operating mode provides clean speech quality equivalent to the ITU-T G.729E codec at 11.8 kbit/s.

The codec operates on 20-ms frames and has a maximum algorithmic delay of 42.875 ms for wideband input and wideband output signals. The maximum algorithmic delay for narrow-band input and narrow-band output signals is 43.875 ms. The codec may also be employed in a low-delay mode when the encoder and decoder maximum bit rates are set to 8 kbit/s or 12 kbit/s. In this case, the maximum algorithmic delay is reduced by 10 ms.

The codec also incorporates an alternate coding mode, with a minimum bit rate of 12.65 kbit/s, which is bitstream interoperable with Recommendation ITU-T G.722.2, 3GPP AMR-WB and 3GPP2 VMR-WB mobile WB speech coding standards. This option replaces layer 1 and layer 2, and the layers 3-5 are similar to the default option with the exception that in layer 3 fewer bits are used to compensate for the extra bits of the 12.65 kbit/s core. The decoder is further able to decode all other ITU-T G.722.2 operating modes. Furthermore, a new annex to this Recommendation is under



development that will efficiently enable bit-stream interoperability with the 3GPP2 EVRC-WB codec. This Recommendation also includes discontinuous transmission mode (DTX) and comfort noise generation (CNG) algorithms that enable bandwidth savings during inactive periods. An integrated noise reduction algorithm can be used provided that the communication session is limited to 12 kbit/s.

The underlying algorithm is based on a two-stage coding structure: the lower two layers are based on code-excited linear prediction (CELP) coding of the band (50-6400 Hz) where the core layer takes advantage of signal classification to use optimized coding modes for each frame. The higher layers encode the weighted error signal from the lower layers using overlap-add modified discrete cosine transformation (MDCT) transform coding. Several technologies are used to encode the MDCT coefficients to maximize performance for both speech and music.

Corrigendum 1 (11/2008) corrects a number of minor problems that have been identified in the fixed-point ANSI C source code of the base text of this Recommendation.

Amendment 1 (03/2009) introduces some additional minor corrections to the fixed-point ANSI C source code and to the text of the Recommendation. It also describes an addition of a verification of the default value of the layer 5 unused bit, and the procedure of erasure of layer 5 if the bit does not have the default value. Amendment 1 also introduces the new Annex A, which defines an alternative implementation of the ITU-T G.718 algorithm using floating point arithmetic to be used for implementation on DSP hardware optimized for floating-point operations. The accompanying floating point ANSI C source code is fully interoperable with the fixed-point code.

While Corrigendum 2 (08/2009) includes further corrections to address minor problems found in both the fixed and floating-point implementations, its main benefit is in the streamlining of the fixed-point implementation which reduces the complexity of the codec from 69 to 57 WMOPS whilst remaining bit-exact with the original code on both steps of the characterization text. This 17% complexity reduction is significant and will clearly make the G.718 more attractive to implement.

Amendment 2 (03/2010) corrects minor defects identified in Recommendation ITU-T G.718 main body (text and ANSI C source code) and introduces new Annex B on a scalable superwideband (50-14000 Hz) extension for ITU-T G.718 operating from 36 to 48 kbit/s.

Corrigendum 3 (01/2011) includes corrections to address minor problems found in both the fixed and floating-point implementations. It also corrects some small inaccuracies found in the Recommendation ITU-T G.718 text and revises the test vectors package to improve the coverage. Corrigendum 3 includes an enhancement designed to improve the performance of music with high frequency pitches beyond 381 Hz (16000 / 42) coded at 16, 24 and 32 kbit/s. The enhancement is a decoder-only option, hence it does not affect ITU-T G.718 bitstream.

This Recommendation contains an electronic attachment with the ANSI C source code, which is an integral part of this Recommendation. Amendment 2 introduces Version 1.5 of the ANSI-C code. Test vectors for the superwideband extension of Annex B are also available online in the ITU-T test signal database at <http://itu.int/net/ITU-T/sigdb/speaudio/Gseries.htm#G.718>. For consistency, the C source code of the complete distribution of Corrigendum 3 (03/2010) has been denoted as "Software Release 1.6".

History

Edition	Recommendation	Approval	Study Group
1.0	ITU-T G.718	2008-06-13	16
1.1	ITU-T G.718 (2008) Cor.1	2008-11-13	16
1.2	ITU-T G.718 (2008) Amend. 1	2009-03-16	16
1.3	ITU-T G.718 (2008) Cor.2	2009-08-29	16
1.4	ITU-T G.718 (2008) Amend.2	2010-03-29	16
1.5	ITU-T G.718 (2008) Cor.3	2011-01-13	16

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2011

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

Page

5.5	ITU-T G.722.2-interoperable option	1
6.13	ITU-T G.722.2-interoperable option	16
7.1	Core layer decoding (layer 1)	17
7.13	Decoding in ITU-T G.722.2-interoperable option	21
8	Description of the transmitted parameter indices	23
8.1	Bit allocation for the default option.....	23
9.2	Organization of the simulation software	24
Annex A	Reference floating-point implementation for ITU-T G.718.....	26
A.5	ANSI C-code	26
Annex B	Superwideband scalable extension for ITU-T G.718	27
Bibliography		27

Electronic attachment: ANSI C source code and test vectors

Recommendation ITU-T G.718

Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s¹

Corrigendum 3

Corrections to text and C-code

Modifications introduced by this corrigendum are shown in revision marks. Unchanged text is replaced by ellipsis (...). Some parts of unchanged text (clause numbers, etc.) may be kept to indicate the correct insertion points.

...

5.5 ITU-T G.722.2-interoperable option

To satisfy the objective of interoperability with other standards, ITU-T G.718 is equipped with an option to allow it to interoperate with ITU-T G.722.2 at 12.65 kbit/s. When invoked, the option allows ITU-T G.722.2 mode 2 (12.65 kbit/s) to replace layer 1 and layer 2. Note that this feature makes the codec interoperable also with the 12.65 kbit/s mode 2 of [29] and mode 3 of [2425]. The decoder is further able to decode all ITU-T G.722.2/AMR-WB coding modes.

...

6.1.5 Frame energy

The spectral analysis module also calculates several energy-related parameters. For example, an average energy in critical bands is computed as

$$E_{CB}(i) = \frac{1}{(L_{FFT}/2)^2 M_{CB}(i)} \sum_{k=0}^{M_{CB}(i)-1} (X_R^2(k+j_i) + X_I^2(k+j_i)), \quad i = 0, \dots, 19 \quad (11)$$

where $X_R(k)$ and $X_I(k)$ are, respectively, the real and imaginary parts of the k -th frequency bin and j_i is the index of the first bin in the i -th critical band given by $j_i = \{1, 3, 5, 7, 9, 11, 13, 16, 19, 22, 26, 30, 35, 41, 47, 55, 64, 75, 89, 107\}$.

Furthermore, energy per frequency bin, $E_{BIN}(k)$, is calculated as

$$E_{BIN}(k) = \frac{1}{(L_{FFT}/2)^2} (X_R^2(k) + X_I^2(k)), \quad k = 0, \dots, 127 \quad (12)$$

$$\overline{E_{BIN}(k) = X_R^2(k) + X_I^2(k)}, \quad k = 0, \dots, 127$$

¹ This Recommendation contains an electronic attachment with the ANSI C source code, which is an integral part of this Recommendation.

...

6.4.3 Levinson-Durbin algorithm

The modified autocorrelation function, $r'(k)$, is used to obtain the LP filter coefficients a_k , $k = 1, \dots, 16$, by solving the set of equations:

$$\sum_{k=1}^{16} a_k r'(|i-k|) = -r'(i), \quad i = 1, \dots, 16 \quad (50)$$

The set of equations (50) is solved using the Levinson-Durbin algorithm [15]. This algorithm uses the following recursion:

$$E(0) = r'(0)$$

for $i = 1$ to 16

$$k_i = -\left[r'(i) + \sum_{j=1}^{i-1} a_j^{(i-1)} r'(i-j) \right] / E(i-1) \quad \underline{k_i = -\left[r'(i) + \sum_{j=1}^{i-1} a_j^{(i-1)} r'(i-j) \right] / E(i-1)}$$

$$a_i^{(i)} = k_i \quad (51)$$

for $j = 1$ to $i-1$

$$a_j^{(i)} = a_j^{(i-1)} + k_i a_{i-j}^{(i-1)}$$

$$E(i) = (1 - k_i^2) E(i-1)$$

The final solution is given as $a_j = a_j^{(16)}$, $j = 1, \dots, 16$, and $a_0 = 1$.

6.4.4 LP to ISP conversion

The LP filter coefficients a_k , $k = 1, \dots, 16$, are converted to the ISP and ISF representation [5] for quantization and interpolation purposes. For a 16th-order LP filter, the ISPs are defined as the roots of the sum and difference polynomials

$$\underline{F_1'(z) = A(z) + z^{-16} A(z^{-1})} \quad \underline{f_1'(z) = A(z) + z^{-16} A(z^{-1})}, \quad (52)$$

$$\underline{F_2'(z) = A(z) - z^{-16} A(z^{-1})} \quad \underline{f_2'(z) = A(z) - z^{-16} A(z^{-1})}. \quad (53)$$

The polynomials $\underline{F_1'(z)}$ and $\underline{F_2'(z)}$ are symmetric and asymmetric, respectively. It can be proven that all roots of these polynomials lie on the unit circle and are interlaced. Polynomial $\underline{F_2'(z)}$ has two roots at $z = 1$ ($\omega = 0$) and $z = -1$ ($\omega = \pi$). To eliminate these two roots, we define the new polynomials

$$\underline{F_1(z) = F_1'(z)} \quad \underline{f_1(z) = f_1'(z)}, \quad (54)$$

$$\underline{F_2(z) = F_2'(z) / (1 - z^{-2})} \quad \underline{f_2(z) = f_2'(z) / (1 - z^{-2})}. \quad (55)$$

Polynomials $\underline{F_1(z)}$ and $\underline{F_2(z)}$ have 8 and 7 conjugate roots on the unit circle ($e^{\pm j\omega_i}$), respectively. Therefore, the polynomials can be written as

$$F_1(z) = (1 + a_{16}) \prod_{i=0,2,\dots,14} (1 - 2q_i z^{-1} + z^{-2}), \quad (56)$$

$$F_2(z) = (1 - a_{16}) \prod_{i=1,3,\dots,13} (1 - 2q_i z^{-1} + z^{-2}) \quad F_2(z) = (1 + a_{16}) \prod_{i=1,3,\dots,13} (1 - 2q_i z^{-1} + z^{-2}), \quad (57)$$

where $q_i = \cos(\omega_i)$ with ω_i being the immittance spectral frequencies (ISFs). The immittance spectral parameters consist of q_i , for $i = 0, \dots, 14$, and q_{15} , which is set to the last prediction coefficient, that is, $q_{15} = a_{16}$ and a_{16} is the last predictor coefficient. The ISFs satisfy the ordering property $0 < \omega_0 < \omega_1 < \dots < \omega_{14} < \pi$. We refer to q_i as the ISPs (in the cosine domain) and ω_i as the ISFs (in the frequency domain).

Since both polynomials $F_1(z)$ and $F_2(z)$ are symmetric, only the first 8 and 7 coefficients of each polynomial, respectively, ~~and the last predictor coefficient~~ need to be computed. The coefficients of these polynomials are found by the following recursive relations:

$$f_1(8) = 2a_8$$

$$\text{for } i = 0 \text{ to } 7 \quad (58)$$

$$f_1(i) = a_i + a_{m-i}$$

$$f_2(i) = a_i - a_{m-i} + f_2(i-2),$$

where $m = 16$ is the predictor order, and $f_2(-2) = f_2(-1) = 0$.

The ISPs are found by evaluating the polynomials $F_1(z)$ and $F_2(z)$ at 100 points equally spaced between 0 and π and checking for sign changes. A sign change indicates the existence of a root and the sign change interval is then divided four times to track the root precisely. The Chebyshev polynomials are used to evaluate $F_1(z)$ and $F_2(z)$ [16]. In this method, the roots are found directly in the cosine domain $\{q_i\}$. The polynomials $F_1(z)$ and $F_2(z)$, evaluated at $z = e^{j\omega}$, can be written as

$$F_1(\omega) = 2e^{-j8\omega} C_1(x),$$

$$F_2(\omega) = 2e^{-j7\omega} C_2(x), \quad (59)$$

with

$$C_1(x) = \sum_{i=0}^7 f_1(i) T_{8-i}(x) + f_1(8)/2, \quad C_1(x) = \sum_{i=0}^7 f_1(i) T_{8-i}(x) + f_1(8)/2,$$

$$C_2(x) = \sum_{i=0}^6 f_2(i) T_{7-i}(x) + f_2(7)/2, \quad C_2(x) = \sum_{i=0}^6 f_2(i) T_{8-i}(x) + f_2(7)/2, \quad (60)$$

where $T_m = \cos(m\omega)$ is the m -th order Chebyshev polynomial and ~~$f(i)$~~ $f_1(i)$ and $f_2(i)$ are the coefficients of ~~either $F_1(z)$ or $F_2(z)$~~ and $F_2(z)$, respectively, computed using equations (58). The polynomial $C(x)$ (either $C_1(x)$ or $C_2(x)$) is evaluated at a certain value of $x = \cos(\omega)$ using the recursive expression:

for $k = n_f - 1$ down to 1

$$b_k = 2xb_{k+1} - b_{k+2} + f(n_f - k) \quad (61)$$

$$C(x) = xb_1 - b_2 + f(n_f)/2,$$

$$C(x) = xb_1 - b_2 + f(n_f)/2$$

where $f(i)$ are coefficients corresponding to either $F_1(z)$ or $F_2(z)$. Further, $n_f = 8$ in case of $C_1(x)$ and $n_f = 7$ in case of $C_2(x)$, with initial values $b_{n_f} = f(0)$ and $b_{n_f+1} = 0$. The detail of the Chebyshev polynomial evaluation method can be found in [16].

6.4.5 ISP to LP conversion

The ISPs, calculated in the previous section, are quantized (see clause 6.8.2 for more details) and interpolated (see clause 6.8.3 for more details). The quantized and unquantized ISPs are converted back to the LP coefficient domain $\{a_k\}$. The conversion is derived below for unquantized values (the conversion of quantized values is done using the same steps). The conversion to the LP domain is done as follows. The coefficients of polynomials $F_1(z)$ and $F_2(z)$ are found by expanding equations (56) and (57), knowing the unquantized and interpolated ISPs q_i , $i = 0, \dots, m-1$, where $m = 16$. The following recursive relation is used to compute coefficients $f_1(z)$

for $i = 2$ to $m/2$

$$\begin{aligned}
 &— f_1(i) = -2q_{2i-2}f_1(i-1) + 2f_1(i-2) \\
 &— \text{for } j = i-1 \text{ down to } 2 \\
 &— f_1(j) = f_1(j) - 2q_{2i-2}f_1(j-1) + f_1(j-2) \\
 &— f_1(1) = f_1(1) - 2q_{2i-2}
 \end{aligned} \tag{62}$$

with initial values $f_1(0) = 1$ and $f_1(1) = -2q_0$. The coefficients $f_2(i)$ are computed similarly, by replacing q_{2i-2} by q_{2i-1} and $m/2$ by $m/2-1$, and with initial conditions $f_2(0) = 1$ and $f_2(1) = -2q_1$.

Once the coefficients $f_1(z)$ and $f_2(z)$ are found, they are multiplied by $(1 + q_{m-1})$ and $(1 - q_{m-1})$, respectively, to form polynomials $F_1(z)$ and $F_2(z)$. $F_2(z)$ is multiplied by $1 - z^{-2}$ to obtain $F'_2(z)$. That is

$$\begin{aligned}
 f_1(i) &= (1 + q_{m-1})f_1(i), & i = 0, \dots, m/2. \\
 f_2(i) &= (1 - q_{m-1})f_2(i), & i = 0, \dots, m/2 - 1 \\
 f'_2(i) &= f_2(i) - f_2(i-2), & i = 2, \dots, m/2 - 1 \\
 f'_1(i) &= f_1(i), & i = 0, \dots, m/2.
 \end{aligned} \tag{63}$$

Then polynomial $F_2(z)$ is multiplied by $(1 - z^{-2})$ to obtain the polynomial $F'_2(z)$. That is coefficients of polynomials $F'_1(z)$ and $F'_2(z)$ are obtained as: multiplied by $1 + q_{m-1}$ and $1 - q_{m-1}$, respectively. That is

$$\begin{aligned}
 f'_1(i) &= f_1(i), & i = 0, \dots, m/2. \\
 f'_2(i) &= f_2(i) - f_2(i-2), & i = 2, \dots, m/2 - 1 \\
 f'_2(i) &= (1 - q_{m-1})f'_2(i), & i = 2, \dots, m/2 - 1 \\
 f'_1(i) &= (1 + q_{m-1})f'_1(i), & i = 0, \dots, m/2.
 \end{aligned} \tag{64}$$

Finally, the LP coefficients are found by

$$\begin{aligned}
a_i &= 0.5f'_1(i) + 0.5f'_2(i), & i &= 1, \dots, m/2 - 1 \\
&0.5f'_1(i) - 0.5f'_2(i), & i &= m/2 + 1, \dots, m - 1 \\
&0.5f'_1(m/2), & i &= m/2 \\
q_{m-1}, & & i &= m
\end{aligned} \tag{65}$$

This is directly derived from the equation $A(z) = (F'_1(z) + F'_2(z))/2$, and considering the fact that $F'_1(z)$ and $F'_2(z)$ are symmetric and asymmetric polynomials, respectively.

...

6.6.3 Normalized correlation computation

After weighting the correlation function with the triangular window of equation (70) centred at the extrapolated pitch lag, the maxima of the weighted correlation function in each of the four sections (three sections, if the first section is not used) are determined. This is performed for both pitch delay ranges. Note that the first section is used only during stable high-pitched speech segments, i.e., if the open-loop pitch period of the second half-frame of the previous frame is lower than or equal to 24 and the scaling factor α_{pn} is higher than or equal to 0.1. In the following, the special case of three sections will not be explicitly dealt with if it arises directly from the text. The pitch delays that yield the maximum of the weighted correlation function will be denoted as $d_{\max}(k)$, where $k = 0, 1, 2, 3$ denotes each of the four sections. Then, the original (raw) correlation function at these pitch delays (pitch lags) is normalized as

$$\begin{aligned}
C_{\text{norm}}(d_{\max}(k)) &= \frac{C(d_{\max}(k))}{\sqrt{\sum_{n=0}^{L_{\text{sec}}} s_{hd}^2(n) \sum_{n=0}^{L_{\text{sec}}} s_{hd}^2(n - d_{\max}(k))}}, \\
C_{\text{norm}}(d_{\max}(k)) &= \frac{C(d_{\max}(k))}{\sqrt{\sum_{n=0}^{L_{\text{sec}}} s_h^2(n) \sum_{n=0}^{L_{\text{sec}}} s_h^2(n - d_{\max}(k))}}, \quad k = 0, 1, 2, 3
\end{aligned} \tag{73}$$

The same normalization is applied also to the weighted correlation function, $C_w(d)$, which yields $C_{wn}(d)$.

...

6.6.4 Correlation reinforcement with pitch lag multiples

In order to avoid selecting pitch multiples within each pitch delay range, the weighted normalized correlation in a lower pitch delay section is further emphasized if one of its multiples is in the neighbourhood of the pitch lag in a higher section. That is,

$$\text{if } \left(|k \times d_{\max}(2) - d_{\max}(3)| \leq k \right)$$

$$\text{if } (\alpha_{\text{prev}} < 0.6 \quad \vee \quad d_{\max}(2) > 0.4 d_{\text{prev}})$$

$$C_{wn}(d_{\max}(2)) = \alpha_{\text{mult}} C_{wn}(d_{\max}(2))$$

$$\alpha_{mult} = (\alpha_{mult})^2$$

$$\text{if } \left(|k \times d_{\max}(1) - d_{\max}(2)| \leq k \right)$$

$$\text{if } \left(\alpha_{prev} < 0.6 \quad \vee \quad d_{\max}(1) > 0.4 d_{prev} \right)$$

$$C_{wn}(d_{\max}(1)) = \alpha_{mult} C_{wn}(d_{\max}(1)),$$

where $\alpha_{mult} = 1.17$, α_{prev} is a voicing factor (normalized pitch correlation) from the previous frame, and d_{prev} is the pitch value from the second half-frame of the previous frame. Further, α_{prev} is given by the voicing factor of the second half-frame in the previous frame if the normalized correlation in the second half-frame was stronger than in the first half-frame, or otherwise by the mean value of these two normalized correlations. In this way, if a pitch period multiple is found in a higher section, the maximum weighted correlation in the lower section is emphasized by a factor of 1.17. However, if the pitch lag in section 3 is a multiple of the pitch lag in section 2 and at the same time the pitch lag in section 2 is a multiple of the pitch lag in section 1, the maximum weighted correlation in section 1 is emphasized twice. This correlation reinforcement is, however, not applied at each section when the previous frame voicing factor, α_{prev} , is more less than or equal to 0.6 and the pitch value is less than or equal to 0.4 times the previous pitch value (i.e., the pitch value does not appear to be a halved value of the previous frame pitch or larger). In this way, the emphasis of the correlation value is allowed only if the correlation in previous frame was low during clear voicing conditions or when the value can be considered to belong to the past pitch contour.

...

6.6.5 Initial pitch lag determination and reinforcement based on pitch coherence with other half-frames

An initial set of pitch lags is determined by searching for the maximum weighted normalized correlation in the four sections in each of the three half-frames. This is done independently for both pitch delay ranges. The initial set of pitch lags is given by

$$d_{init}^{[k]} = \arg \max_{i=0}^3 \left(C_{wn}(d_{\max}^{[k]}(i)) \right), \quad \text{for } k = 0, 1, 2,$$

where the superscript $[k]$ denotes the first, the second or the third (look-ahead) half-frame.

To find the right pitch value, another level of weighting is performed on the weighted normalized correlation function, $C_{wn}(d)$, in each section of each half-frame in each pitch delay range. This weighting is based on pitch coherence of the initial set of pitch lags, $d_{init}^{[i]}$, with pitch lags $d_{\max}^{[j]}(k)$ ~~$d_{init}^{[j]}(k)$~~ , $k = 0, 1, 2, 3, j \neq i$, i.e., those from the other half-frames. The weighting is further reinforced with pitch lags selected from the complementary pitch delay range, denoted as $d_{\max}^{[j]}(k)$ ~~$d_{init}^{[j]}(k)$~~ , $k = 0, 1, 2, 3, j \neq i$. Further, the weighting favours section-wise stability, where a stronger weighting is applied for coherent pitch values that are from the same section of the same set as the initial pitch lag, and a slightly weaker weighting is applied for coherent pitch values that are from a different section and/or a different pitch delay range than the initial pitch lag. That is, if the initial pitch lag in a half-frame i is coherent with a pitch lag of section k in half-frame j , then the corresponding