FACULTY OF ELECTRICAL ENGINEERING
AND INFORMATION TECHNOLOGY
DEPARTMENT OF COMMUNICATIONS ENGINEERING

TECHNICAL REPORT

Derivation of the Power Compensation Constant in the Observation Model for Reverberant Speech in the Logarithmic Mel Power Spectral Domain

December 20, 2012

Authors:

Volker Leutnant
leutnant@nt.uni-paderborn.de

Alexander Krueger
krueger@nt.uni-paderborn.de

Reinhold Haeb-Umbach
haeb@nt.uni-paderborn.de

CONTENTS

The power compensation constant plays an important role in the observation model for reverberant speech signals [1], [2], whose derivation will be repeated here for convenience.

Note that to introduce the used notation and to set this technical report into proper context, Sec. I and Sec. II are directly taken from [1], which has recently been submitted to *IEEE Transactions on Audio, Speech And Language*. The experienced reader may thus skip these sections and directly proceed to Sec. III.

I. Observation model for reverberant-only speech signals

In the absence of background noise, the discrete-time microphone signal $\overline{s}(l)$ is given by the convolution of the clean speech signal $\overline{x}(l)$ with the acoustic impulse response (AIR) $h(l)$. According to [2], the corresponding STDFT $S(m,k)$ may be expressed as

$$S(m,k) = \sum_{k'=0}^{K-1} \sum_{m'=-L_{H,\ell}}^{L_H} X(m-m',k') h_{k,k'}(m') \tag{1}$$

with

$$h_{k,k'}(m') := \sum_{p'=0}^{L_h-1} h(p') \phi_{k,k'}(m'B - p') \tag{2}$$

and

$$\phi_{k,k'}(l) := e^{j\frac{2\pi}{K}k'l} \sum_{l'=0}^{L_w-1} w_{\mathrm{A}}(l') w_{\mathrm{S}}(l'+l) e^{-j\frac{2\pi}{K}(k-k')l'}. \tag{3}$$

The terms $h_{k,k'}(m')$ will in the following be referred to as cross-band filters for $k \neq k'$ and as band-to-band filters for $k = k'$, as in [3]. The lengths $L_{H,\ell}$ and L_H in (1) are defined by

$$L_{H,\ell} := \left\lfloor \frac{L_w - 1}{B} \right\rfloor, \quad L_H := \left\lfloor \frac{L_h + L_w - 2}{B} \right\rfloor. \tag{4}$$

Further, $w_{\mathrm{S}}(l')$ denotes a synthesis window, which is bi-orthogonal to $w_{\mathrm{A}}(l')$ [4] and has the same support as $w_{\mathrm{A}}(l')$. For the power of $S(m,k)$ we now write

$$|S(m,k)|^2 = C_{\mathrm{P}} \sum_{m'=0}^{L_H} \left|X(m-m',k)\right|^2 \left|h_{k,k}(m')\right|^2 + E^{(S)}(m,k). \tag{5}$$

The introduced constant C_{P} and the error term $E^{(S)}(m,k)$ thereby capture all terms incurring when the square of the sum given in (1) is approximated by the sum of the squares while also ensuring a causal relationship by dropping all negative frame indices. The constant C_{P} will be determined such that the error term $E^{(S)}(m,k)$ is zero-mean, or, equivalently,

$$\mathrm{E}\left[\left|\breve{S}(m,k)\right|^2\right] \stackrel{!}{=} \mathrm{E}\left[C_{\mathrm{P}} \sum_{m'=0}^{L_H} \left|\breve{X}(m-m',k)\right|^2 \left|\breve{h}_{k,k}(m')\right|^2\right]. \tag{6}$$

Note that in (6) and in the following, we use the breve mark $(\breve{\cdot})$ to distinguish a random variable from its realization. Due to the role of the constant C_{P} in (6), it is in the following referred to as the power compensation constant. Besides being additive in the targeted LMPSC domain, choosing a multiplicative constant C_{P} rather than an additive term to compensate for the bias introduced by the approximation is advantageous, since the desired compensation is made independent of the power of the clean speech signal and that of the AIR.

Note, that the approximation presented here is more general than that in [2], where we chose an empirical value of $(L_w/B)^2$.

By further introducing the mean of $|h_{k,k}(m')|^2$ over the qth mel band, i.e.,

$$\bar{\mathcal{H}}_{m',q} := \frac{1}{K_q^{(\mathrm{u})} - K_q^{(\ell)} + 1} \sum_{k=K_q^{(\ell)}}^{K_q^{(\mathrm{u})}} \left|h_{k,k}(m')\right|^2, \tag{7}$$

the MPSCs $\mathcal{S}_{m,q}$ of the reverberant speech signal can be written as

$$\mathcal{S}_{m,q} = C_{\mathrm{P}} \sum_{m'=0}^{L_H} \bar{\mathcal{H}}_{m',q} \mathcal{X}_{m-m',q} + \mathcal{E}_{m,q}^{(\mathcal{S})} \tag{8}$$

$$=: \tilde{\mathcal{S}}_{m,q} + \mathcal{E}_{m,q}^{(\mathcal{S})}. \tag{9}$$

Hereby $\mathcal{X}_{m,q}$ denote the MPSCs of the clean speech signal and $\mathcal{E}_{m,q}^{(\mathcal{S})}$ the error resulting from the approximation of $\mathcal{S}_{m,q}$ by $\tilde{\mathcal{S}}_{m,q}$.

By introducing the logarithmic mel power spectral representation of the AIR

$$\bar{h}_{m',q} := \ln\left\{\bar{\mathcal{H}}_{m',q}\right\} \tag{10}$$

and the LMPSC of the clean speech signal

$$x_{m,q} := \ln\left\{\mathcal{X}_{m,q}\right\}, \tag{11}$$

we are now able to express the LMPSCs of the reverberant speech in terms of the underlying LMPSCs of the clean speech and the LMPSCs of the AIR, i.e.,

$$s_{m,q} := \ln\left\{\mathcal{S}_{m,q}\right\} = \tilde{s}_{m,q} + v_{m,q}^{(s)}, \tag{12}$$

where

$$\tilde{s}_{m,q} := \ln\left\{\tilde{\mathcal{S}}_{m,q}\right\} = \ln\left\{C_{\mathrm{P}} \sum_{m'=0}^{L_H} e^{x_{m-m',q} + \bar{h}_{m',q}}\right\}. \tag{13}$$

Thereby, employing the definition of $\tilde{\mathcal{S}}_{m,q}$ given in (9) instead of the equivalent formulation given in (13), the additive observation error in the LMPSC domain

$$v_{m,q}^{(s)} := s_{m,q} - \tilde{s}_{m,q} = \ln\left\{\mathcal{S}_{m,q}\right\} - \ln\left\{\tilde{\mathcal{S}}_{m,q}\right\} \tag{14}$$

$$= \ln\left\{\frac{\mathcal{S}_{m,q}}{\sum\limits_{m'=0}^{L_H} \bar{\mathcal{H}}_{m',q}\mathcal{X}_{m-m',q}}\right\} - \ln\left\{C_{\mathrm{P}}\right\} \tag{15}$$

captures the errors from the approximation of $\mathcal{S}_{m,q}$ by $\tilde{\mathcal{S}}_{m,q}$ in the MPSC domain. Note that the choice of C_{P} only affects the mean of the observation error $v_{m,q}^{(s)}$.

By introducing the observation mapping

$$f_s\left(\mathbf{x}_{m-L_H:m}, \bar{\mathbf{h}}_{0:L_H}\right) := \ln\left\{C_{\mathrm{P}} \sum_{m'=0}^{L_H} e^{\mathbf{x}_{m-m'} + \bar{\mathbf{h}}_{m'}}\right\}, \tag{16}$$

where the mathematical operations are understood to be performed on the vectors component-wise, the relationship (12) may compactly be formulated by

$$\mathbf{s}_m = f_s\left(\mathbf{x}_{m-L_H:m}, \bar{\mathbf{h}}_{0:L_H}\right) + \mathbf{v}_m^{(s)}, \tag{17}$$

where $\mathbf{x}_{m-L_H:m} := \{\mathbf{x}_{m-L_H}, ..., \mathbf{x}_m\}$ and $\bar{\mathbf{h}}_{0:L_H} := \left\{\bar{\mathbf{h}}_0, ..., \bar{\mathbf{h}}_{L_H}\right\}$ denote sequences of $L_H + 1$ Q-dimensional vectors.

II. Acoustic impulse response model

The observation model (17) requires the logarithmic mel power spectral representation of the AIR $\bar{\mathbf{h}}_{0:L_H}$. In practice, however, this representation is usually unknown. To avoid a sensitive blind estimation of the AIR for a computation of $\bar{\mathbf{h}}_{0:L_H}$, we have proposed in [2] to employ a stochastic AIR model, which has previously been introduced in [5]. According to this model, the AIR is regarded to be a realization of a stochastic process $\breve{h}(l)$ according to

$$\breve{h}(l) = \sigma_h \breve{v}_h(l) \chi_h(l)\, e^{-\frac{l}{\tau_h}}, \tag{18}$$

where $\breve{v}_h(l)$ is a zero-mean white Gaussian stochastic process of unit power. The indicator function

$$\chi_h(l) := \begin{cases} 1 & \text{for } 0 \leq l \leq L_h - 1 \\ 0 & \text{else} \end{cases} \tag{19}$$

assures the AIR to be causal having a finite length L_h. The term $e^{-\frac{l}{\tau_h}}$ causes an exponentially decaying envelope, where the decay constant τ_h depends on the reverberation time T_{60} through

$$\tau_h = \frac{T_{60}}{3\ln(10)\,T_{\mathrm{S}}}, \tag{20}$$

with T_{S} denoting the sampling duration. The constant σ_h may be used to control the AIR energy. The advantage of using this model is that it has only two parameters, i.e., τ_h and σ_h, which can be estimated more easily than the complete AIR.

Based on the AIR model (18) a reasonable length L_h may be determined in dependence on τ_h by

$$L_h = L_h(\tau_h) = \left\lceil -\frac{\tau_h}{2}\ln(\epsilon_h) \right\rceil, \tag{21}$$

which is obtained by minimizing the AIR length under the constraint that the relative energy of the neglected part of the AIR is smaller than ϵ_h [2].

It has been shown in [6] that the PDFs of the individual components $\breve{\bar{h}}_{m',q}$ of the logarithmic mel power spectral AIR representation $\breve{\bar{\mathbf{h}}}_{0:L_H}$ can be well modeled by GAUSSIANS.

Moreover, their respective means and variances have been found to only depend on the decay constant τ_h, the energy term σ_h and on parameters of the ETSI Standard Front-End.

As we have previously done in [2], we therefore propose to approximate the usually unknown logarithmic mel power spectral representation of the AIR $\breve{\bar{\mathbf{h}}}_{0:L_H}$ by its mean $\boldsymbol{\mu}_{\breve{\bar{\mathbf{h}}}_{0:L_H}}$ under the AIR model (18). By doing so, the time-variance of the AIR due to, e.g., small movements of the speaker is absorbed through the observation error to a certain degree.

Having introduced the stochastic AIR model, we are able to compute the power compensation constant C_{P} from condition (6) by using the AIR model and two additional assumptions. We assume the AIR and the clean speech signal to be mutually independent and the latter to be a realization of a zero-mean white GAUSSIAN stochastic process. Under these assumptions, the derived C_{P} is given by

$$C_{\mathrm{P}} = \frac{C_{\mathrm{N}}}{C_{\mathrm{D}}}, \tag{22}$$

where

$$C_{\mathrm{N}} := K^2 \sum_{m',m''=-L_{H,\ell}}^{L_H} \sum_{l=0}^{L_w-1} w_{\mathrm{A}}(l) w_{\mathrm{S}}(l) w_{\mathrm{A}}(l+\left(m''-m'\right)B) w_{\mathrm{S}}(l+\left(m''-m'\right)B) \cdot \sum_{l'=-L_w+1}^{L_w-1} \chi_h\left(m'B-l'\right) e^{-\frac{2\left(m'B-l'\right)}{\tau_h}} w_{\mathrm{A}}^2(-l'+l), \tag{23}$$

$$C_{\mathrm{D}} := \left[\sum_{m'=0}^{L_H} \sum_{l'=-L_w+1}^{L_w-1} w^2(-l')\chi_h\left(m'B-l'\right) e^{-\frac{2\left(m'B-l'\right)}{\tau_h}}\right] \left(\sum_{l=0}^{L_w-1} w_{\mathrm{A}}^2(l)\right) \tag{24}$$

and

$$w(l) := \sum_{l'=0}^{L_w-1} w_{\mathrm{A}}(l') w_{\mathrm{S}}(l'-l). \tag{25}$$

In practice, it is reasonable to precompute the power compensation constant for a set of predefined reverberation times prior to applying any of the observation models, e.g., to feature enhancement and successive ASR, and choose C_{P} based on an estimate of the reverberation time.

III. Derivation of power compensation constant

The constant C_{P} is introduced to obtain a tractable estimate of the power of the STDFT of the reverberant speech signal. We propose to choose it to satisfy the constraint (6), where the expectation is assumed to be taken over all possible clean speech signals and AIRs, which are possible for fixed constants T_{60} and σ_h^2. The basis for the derivation is the stochastic AIR model (18). With it, the band-to-band filters $h_{k,k}(m)$ may be expressed by

$$h_{k,k}(m) = \sum_{p'=0}^{L_h-1} h(p') w(p'-mB) e^{j\frac{2\pi}{K}k(mB-p')} \tag{26}$$

using (2), (3) and (25). After applying the variable substitution $l = p' - mB$ to (26), we obtain

$$h_{k,k}(m) = \sum_{l=-L_w+1}^{L_w-1} h(l+mB) w(l) e^{-j\frac{2\pi}{K}kl}. \tag{27}$$

Assuming the AIR to be a realization of the stochastic process $\breve{h}(l)$ defined in (18) whose auto-correlation function is given by

$$\mathrm{E}\left[\breve{h}(l)\breve{h}(l')\right] = \sigma_h^2 \delta\left(l-l'\right) \chi_h\left(l\right) e^{-\frac{2l}{\tau_h}}, \tag{28}$$

with $\delta\left(l\right)$ denoting the DIRAC delta function, we can employ (27) and (28) to finally obtain

$$\mathrm{E}\left[\left|\breve{h}_{k,k}(m)\right|^2\right] = \mathrm{E}\left[\left|\sum_{l=-L_w+1}^{L_w-1} \breve{h}(l+mB) w(l) e^{-j\frac{2\pi}{K}kl}\right|^2\right] \tag{29}$$

$$= \sum_{l=-L_w+1}^{L_w-1} \sigma_h^2 \, \chi_h\left(l+mB\right) e^{-\frac{2(l+mB)}{\tau_h}} w^2(l) \tag{30}$$

For a feasible solution we further assume the clean speech signal $x(l)$ to be a realization of a white GAUSSIAN stochastic process $\breve{x}(l)$ with power σ_x^2, which is stochastically independent of the AIR. The auto-correlation function of its STDFT may thus be expressed as

$$\mathrm{E}\left[\breve{X}(m-m',k')\breve{X}^*(m-m'',k'')\right]$$
$$= \sigma_x^2 e^{j\frac{2\pi}{K}k''(m''-m')B} \sum_{l=0}^{L_w-1} w_{\mathrm{A}}(l) w_{\mathrm{A}}(l+\left(m''-m'\right)B) e^{-j\frac{2\pi}{K}(k'-k'')l}. \tag{31}$$

The auto-correlation function of the cross-band filters can be written with (2) and (28) as

$$\mathrm{E}\left[\breve{h}_{k,k'}(m')\breve{h}^*_{k,k''}(m'')\right] = \mathrm{E}\left[\sum_{l,l'=0}^{L_h-1}\breve{h}(l)\breve{h}(l')\phi_{k,k'}(m'B-l)\phi^*_{k,k''}(m''B-l')\right] \tag{32}$$

$$= \sigma_h^2\sum_{l=0}^{L_h-1}\chi_h\left(l\right)\phi_{k,k'}(m'B-l)\phi^*_{k,k''}(m''B-l)e^{-\frac{2l}{\tau_h}}. \tag{33}$$

Exploiting the fact that the support of $\phi_{k,k'}(l)$ is given by $[-L_w+1, L_w-1]$, (33) may be reformulated using the variable substitution $l' = m'B - l$ by

$$\mathrm{E}\left[\breve{h}_{k,k'}(m')\breve{h}^*_{k,k''}(m'')\right]$$
$$= \sigma_h^2\sum_{l'=-L_w+1}^{L_w-1}\phi_{k,k'}(l')\phi^*_{k,k''}(l'+\left(m''-m'\right)B)\chi_h\left(m'B-l'\right)e^{-\frac{2\left(m'B-l'\right)}{\tau_h}}. \tag{34}$$

With (31) and (34), the power compensation constant C_P may now be computed by separately considering the left and right hand side of (6).

a) Left hand side of (6)*:* Making use of the stochastic independence of the clean speech signal and the AIR, the left hand side of (6) may be written as

$$\mathrm{E}\left[\left|\breve{S}(m,k)\right|^2\right]$$
$$= \sum_{m',m''=-L_{H,\ell}}^{L_H}\sum_{k',k''=0}^{K-1}\mathrm{E}\left[\breve{X}(m-m',k')\breve{X}^*(m-m'',k'')\right]\mathrm{E}\left[\breve{h}_{k,k'}(m')\breve{h}^*_{k,k''}(m'')\right]. \tag{35}$$

By plugging the found relationships (31) and (34) into (35), we arrive at

$$\mathrm{E}\left[\left|\breve{S}(m,k)\right|^2\right] = \sigma_x^2\sigma_h^2\sum_{m',m''=-L_{H,\ell}}^{L_H}\sum_{l=0}^{L_w-1}w_\mathrm{A}(l)w_\mathrm{A}(l+\left(m''-m'\right)B)$$
$$\cdot\sum_{l'=-L_w}^{L_w-1}\chi_h\left(m'B-l'\right)e^{-\frac{2\left(m'B-l'\right)}{\tau_h}}\cdot\xi_{m''-m',l,l',k}, \tag{36}$$

with

$$\xi_{m''-m',l,l',k} := \sum_{k',k''=0}^{K-1}\phi_{k,k'}(l')\phi^*_{k,k''}(l'+\left(m''-m'\right)B)\cdot e^{-j\frac{2\pi}{K}[k'l-k''(l+(m''-m')B)]}. \tag{37}$$

By further substituting the definition of $\phi_{k,k'}(l)$ into (37) according to (3), we obtain

$$\begin{aligned}
&\xi_{m''-m',l,l',k} \\
&= \sum_{k',k''=0}^{K-1} \left[\sum_{p'=0}^{L_w-1} w_{\mathrm{A}}(p') w_{\mathrm{S}}(p'+l') e^{j\frac{2\pi}{K}k'(p'+l')} e^{-j\frac{2\pi}{K}kp'} \right] e^{-j\frac{2\pi}{K}[k'l-k''(l+(m''-m')B)]} \\
&\qquad \cdot \left[\sum_{p''=0}^{L_w-1} w_{\mathrm{A}}(p'') w_{\mathrm{S}}(p''+l'+\left(m''-m'\right)B) e^{-j\frac{2\pi}{K}k''(p''+l'+(m''-m')B)} e^{j\frac{2\pi}{K}kp''} \right] \qquad (38) \\
&= \sum_{p''=0}^{L_w-1} w_{\mathrm{A}}(p'') w_{\mathrm{S}}(p''+l'+\left(m''-m'\right)B) e^{j\frac{2\pi}{K}kp''} \sum_{p'=0}^{L_w-1} w_{\mathrm{A}}(p') w_{\mathrm{S}}(p'+l') e^{-j\frac{2\pi}{K}kp'} \psi_{p',p'',l,l'} \qquad (39)
\end{aligned}$$

with

$$\psi_{p',p'',l,l'} := \left[\sum_{k'=0}^{K-1} e^{j\frac{2\pi}{K}k'(p'+l'-l)} \right] \left[\sum_{k''=0}^{K-1} e^{-j\frac{2\pi}{K}k''(p''+l'-l)} \right]. \qquad (40)$$

Considering the sum orthogonality

$$\frac{1}{K} \sum_{k=0}^{K-1} e^{j\frac{2\pi}{K}k\mu} = \sum_{\nu=-\infty}^{\infty} \delta\left(\mu - \nu K\right) \text{ for } \mu \in \mathbb{Z}, K \in \mathbb{N}, \qquad (41)$$

the term $\psi_{p',p'',l,l'}$ can be expressed as

$$\psi_{p',p'',l,l'} = K^2 \left[\sum_{\nu'=-\infty}^{\infty} \delta\left(p'+l'-l-\nu' K\right) \right] \left[\sum_{\nu''=-\infty}^{\infty} \delta\left(p''+l'-l-\nu'' K\right) \right]. \qquad (42)$$

Having in mind the identity

$$\delta\left(l-\mu\right) \delta\left(l-\mu'\right) = \delta\left(l-\mu\right) \delta\left(\mu-\mu'\right) \text{ for } l, \mu, \mu' \in \mathbb{Z}, \qquad (43)$$

we may simplify $\psi_{p',p'',l,l'}$ further by

$$\psi_{p',p'',l,l'} = K^2 \left[\sum_{\nu=-\infty}^{\infty} \delta\left(p'+l'-l-\nu' K\right) \right] \left[\sum_{=-\infty}^{\infty} \delta\left(p''-p'-\left(\nu''-\nu'\right) K\right) \right]. \qquad (44)$$

Since the difference $p''-p'$ lies within the interval $[-L_w+1, L_w-1]$ and $K > L_w$ holds, the argument of the second DIRAC delta function in (44) may become zero only for $\nu'' = \nu'$. For that reason

$$\psi_{p',p'',l,l'} = K^2 \sum_{\nu'=-\infty}^{\infty} \delta\left(p'+l'-l-\nu' K\right) \delta\left(p''-p'\right) \qquad (45)$$

holds. By substituting (45) into (39), it follows that

$$\begin{aligned}
&\xi_{m''-m',l,l',k} \\
&= K^2 \sum_{\nu'=-\infty}^{\infty} \sum_{p'=0}^{L_w-1} w_{\mathrm{A}}(p') w_{\mathrm{S}}(p'+l') \delta\left(p'+l'-l-\nu' K\right) \\
&\qquad \cdot \sum_{p''=0}^{L_w-1} w_{\mathrm{A}}(p'') w_{\mathrm{S}}(p''+l'+\left(m''-m'\right) B) \delta\left(p''-p'\right) e^{-j\frac{2\pi}{K} k(p'-p'')} \qquad (46) \\
&= K^2 \sum_{\nu'=-\infty}^{\infty} \sum_{p'=0}^{L_w-1} w_{\mathrm{A}}^2(p') w_{\mathrm{S}}(p'+l') \delta\left(p'+l'-l-\nu' K\right) w_{\mathrm{S}}(p'+l'+\left(m''-m'\right) B) \qquad (47) \\
&= K^2 \sum_{\nu'=-\infty}^{\infty} w_{\mathrm{A}}^2(-l'+l+\nu' K) w_{\mathrm{S}}(l+\nu' K) w_{\mathrm{S}}(l+\nu' K+\left(m''-m'\right) B), \qquad (48)
\end{aligned}$$

From this expression it can be seen that $\xi_{m''-m',l,l',k}$ does not depend on k. Since for $l \in [-L_w+1, L_w-1]$ the equality $w_{\mathrm{S}}(l+\nu' K)=0 \ \forall \nu' \neq 0$ holds, we have

$$\xi_{m''-m',l,l',k} = K^2 w_{\mathrm{A}}^2(-l'+l) w_{\mathrm{S}}(l) w_{\mathrm{S}}(l+\left(m''-m'\right) B). \qquad (49)$$

By substituting (49) into (36), we obtain

$$\mathrm{E}\left[\left|\breve{S}(m,k)\right|^2\right] = \sigma_x^2 \sigma_h^2 \cdot C_{\mathrm{N}} \qquad (50)$$

with

$$\begin{aligned}
C_{\mathrm{N}} := K^2 &\sum_{m',m''=-L_{H,\ell}}^{L_H} \sum_{l=0}^{L_w-1} w_{\mathrm{A}}(l) w_{\mathrm{S}}(l) w_{\mathrm{A}}(l+\left(m''-m'\right) B) w_{\mathrm{S}}(l+\left(m''-m'\right) B) \\
&\cdot \sum_{l'=-L_w+1}^{L_w-1} \chi_h\left(m' B-l'\right) e^{-\frac{2\left(m' B-l'\right)}{\tau_h}} w_{\mathrm{A}}^2(-l'+l). \qquad (51)
\end{aligned}$$

b) Right hand side of (6)*:* The expression on the right hand side of (6) may be simplified using (31) and (34) to be

$$\begin{aligned}
&\mathrm{E}\left[C_{\mathrm{P}} \cdot \sum_{m'=0}^{L_H} \left|\breve{X}(m-m',k)\right|^2 \left|\breve{h}_{k,k}(m')\right|^2\right] \\
&= C_{\mathrm{P}} \sum_{m'=0}^{L_H} \mathrm{E}\left[\left|\breve{X}(m-m',k)\right|^2\right] \mathrm{E}\left[\left|\breve{h}_{k,k}(m')\right|^2\right] \qquad (52) \\
&= C_{\mathrm{P}} \sum_{m'=0}^{L_H} \left(\sigma_x^2 \sum_{l=0}^{L_w-1} w_{\mathrm{A}}^2(l)\right) \left(\sigma_h^2 \sum_{l'=-L_w+1}^{L_w-1} \left|\phi_{k,k}(l')\right|^2 \chi_h\left(m' B-l'\right) e^{-\frac{2\left(m' B-l'\right)}{\tau_h}}\right) \qquad (53) \\
&= C_{\mathrm{P}} \cdot \sigma_x^2 \sigma_h^2 \cdot C_{\mathrm{D}}. \qquad (54)
\end{aligned}$$

Employing (3) for $k{=}k'$ and (25), C_{D} may be expressed as

$$C_{\mathrm{D}} := \left[\sum_{m'=0}^{L_H}\sum_{l'=-L_w+1}^{L_w-1} w^2(-l')\chi_h\left(m'B-l'\right)e^{-\frac{2\left(m'B-l'\right)}{\tau_h}}\right]\left(\sum_{l=0}^{L_w-1} w_{\mathrm{A}}^2(l)\right). \tag{55}$$

The desired constant C_{P} is finally obtained from comparing expressions (54) and (50) to be

$$C_{\mathrm{P}} = \frac{C_{\mathrm{N}}}{C_{\mathrm{D}}}. \tag{56}$$

From (51) and (55), it can be seen that the constant C_{P} only depends on the parameters employed for feature extraction and on the reverberation time.

The power compensation constant obtained for the parameter values according to the ETSI Standard Front-End [7] are given in Fig. 1 for different reverberation times. The required AIR parameters for different reverberation times resulting from the choice $\epsilon_h = 10^{-3}$ are listed in Tab. I It can be seen that

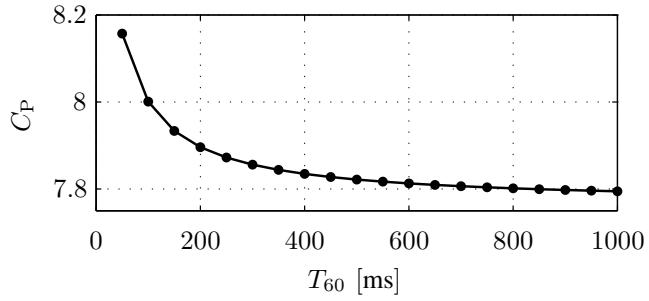

Fig. 1. Power compensation constant for different reverberation times.

the constant depends only on the parameters for the extraction, i.e. the analysis and synthesis windows, the number of frequency bins etc., and on the reverberation time and that its value is about 8 for a large range of practically relevant reverberation times.

In practice, it is reasonable to precompute the power compensation constant for a set of predefined reverberation times prior to applying any of the observation models, e.g., to feature enhancement and successive ASR, and choose C_{P} based on an estimate of the reverberation time.

TABLE I

AIR PARAMETERS FOR DIFFERENT REVERBERATION TIMES RESULTING FROM THE CHOICE $\epsilon_h = 10^{-3}$.

$\hat{\mathbf{T}}_{\mathbf{60}}$	250 ms	350 ms	450 ms	550 ms	650 ms
AIR length $\hat{\mathbf{L}}_{\mathbf{h}}$	1000	1400	1800	2200	2600
LMPSC length $\hat{\mathbf{L}}_{\mathbf{H}}$	14	19	24	29	34

REFERENCES

[1] V. Leutnant, A. Krueger, and R. Haeb-Umbach, "A new observation model in the logarithmic mel power spectral domain for the automatic recognition of noisy reverberant speech," Sep. 2012, submitted for publication in IEEE Trans. Audio, Speech, Lang. Process.

[2] A. Krueger and R. Haeb-Umbach, "Model-based feature enhancement for reverberant speech recognition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1692–1707, 2010.

[3] Y. Avargel and I. Cohen, "System identification in the short-time Fourier transform domain with crossband filtering," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1305–1319, May 2007.

[4] S. Qian and D. Chen, "Discrete Gabor transform," *IEEE Trans. Signal Process.*, vol. 41, no. 7, pp. 2429–2438, Jul. 1993.

[5] J. Polack, "La transmission de l'énergie sonore dans les salles," Dissertation, Université du Maine, 1988.

[6] A. Krüger, "Modellbasierte Merkmalsverbesserung zur robusten automatischen Spracherkennung in Gegenwart von Nachhall und Hintergrundstörungen," Ph.D. dissertation, University of Paderborn, Germany, Dec. 2011.

[7] ETSI, *ETSI standard document, Speech Processing, Transmission and Quality Aspects (STQ); Distributed speech recognition; Front-end feature extraction algorithm; Compression algorithms, ETSI ES 201 108 V1.1.3 (2003-09).*