Responsible Authors: Peter Patrick

Issue: 1.2

SW and Algorithmic Design Description of the

Handsfree Module in G27

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

Issue: 1.2

List of Updates

Issue	Date	Updated Paragraphs	Description
1.0p	17 Feb 1999	all	Creation of document (proposal)
1.1p	01 Mar 1999	all	Thoroughly reworked
1.2	04 Mar 1999	all	Fine tuned and finalised

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

1. General

1.1. Purpose

This document contains the algorithmic description and software organisation of the Handsfree speech package used within the context of the G27 project which is deployed in the baseband IC PCF5087 ref.[2]. The package functions, outlines and implementation aspects as used in the C-reference program respective to the R.E.A.L.-assembler code are also presented. It is intended to describe the principle algorithms applied rather than any details of the fixed-point DSP-implementation.

1.2. Scope

The handsfree package is intended to be used in the GSM hand-held mobile telephone using the IC PCF5087

1.3. Organisation of this Document

The algorithmic description of the handsfree package which is contained in section (2) starts with a principle overview of the handsfree functions and continues to describe the algorithms used to implement the subfunctions in more detail. This corresponds to the order in which they are processed by the DSP-software. This document also encompasses a description of the accompanying C-reference code used as a basis for the simulation platform of the algorithm and used to generate the bit exact test sequences used to align the R.E.A.L. assembler firmware. The topics are ordered as follows:

- Description of the handsfree modules comprising of the weighting balance and comfort noise generation
- Functional block schematic diagram
- IO-Files for the Handsfree unit
- List of routines and respective files
- SW-calling tree

The treatment of the corresponding R.E.A.L. assembler code is not dealt with in this document. For brief usage aspects, the reader is referred to the interface specifications ref.[3] and ref.[7]

1.4. Modification Procedure

The modification procedure of this document is as defined in ref.[1].

1.5. Distribution

This document will be distributed to:

Table 1: Distribution List

Name	***Function***	***Location***
Harald Bauer	GSM-System architect	TCMC, Nürnberg
Volker Budig	G27-Project Leader	TCMC, Nürnberg
Harald Weiss	DSP-Expert/Engineer	TCMC, Nürnberg

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

Table 1: Distribution List

Name	*Function*	*Location*
Rainer Dietsch	DSP-Expert/Engineer	PS-SV, Sunnyvale (secondment from TCMC)
TCMC-Info-System		TCMC-Nürnberg

1.6. Applicability

1.7. Abbreviations

CN	Comfort Noise
DC	Direct Current (zero frequency spectral contribution)
DTX	Discontinuous Transmission
DSP	Digital Signal Processing/Processor
HF	Handsfree Speech Module/Package
LPC	Linear Prediction Coding/Coefficients
Rx	Receive
SW	Software
TCMC	Technology Centre for Mobile Communication
Tx	Transmit
VAD	Voice Activity Detector

1.8. Glossary

1.9. References

[1] P. Hirt: “Document, Numbering, Release, Change”, Version 1.0, ABV09-GEN-VA003

[2] H. Bauer, V. Delong: "PCF508X Multi-Task DSP Firmware", ABD06-G27-DD001

[3] R. Dietsch: “Handsfree Package Interface Description for the R.E.A.L. Digital Signal Processor", ABC05-G26-IS001

[4] Digital Cellular Telecommunications System (Phase 2+) Full Rate Speech Transcoding (GSM 06.10 version 5.1.1) May 1999

[5] Voice Activity Detector (VAD) for Full Rate Speech Traffic Channels (GSM 06.32 version 5.0.3) April 1998

[6] Comfort Noise Aspects for full rate Speech Traffic Channels (GSM-06.12 version 5.0.1) May 1997

[7] H. Weiss: "SW-Interface Specification GSM Speech Coding Package for the R.E.A.L. DSP" ABC08-G27-IS008

2. Algorithmic Description

2.1. General Principles

The use of a hand-held mobile telephone whilst driving a motor vehicle radically impairs traffic road safety due to unnecessary distraction from appropriate care and attention necessary to control the vehicle and to observe road conditions. In the UK, the Highway Code forbids handset usage by drivers unless the motor vehicle is stationary parked. It is, however, not currently a punishable offence in itself. Other countries may inevitably follow suit. Traffic insurance companies view that

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

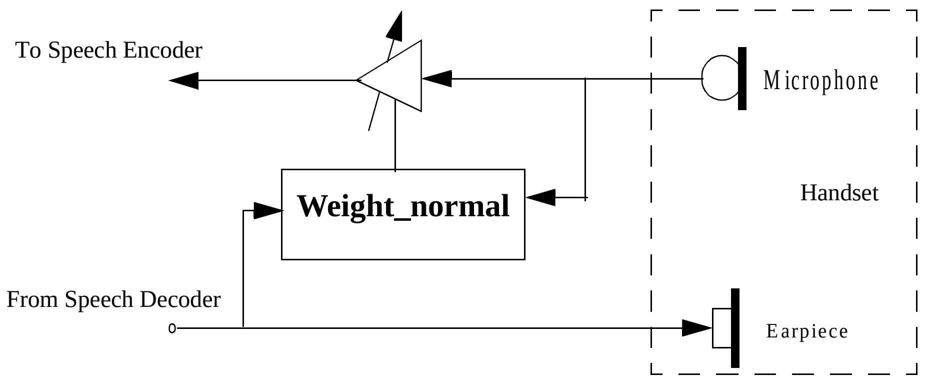
Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

drivers using hand-held mobile telephones being involved in a traffic accident can have part or whole of the liability apportioned to themselves which adversely affect corresponding damage claims.

In order to overcome this problem, the handset may be replaced by a physically fixed and separated microphone and loudspeaker unit (i.e. the *car kit* unit) which, in essence, frees the mobile user of one or both hands (hence the term *handsfree***)**. The inherent acoustic echo effect from the loudspeaker to the microphone must be alleviated by the use of a dynamic voice switch to enable an automatic semi-duplex operation. Hard switching the loudspeaker and microphone in dependency of speech activity is very unpleasant for the users, especially if one or both parties are in acoustically noisy environments. If the hard switch is replaced by smoothed or a gradual switch, then this can be used as the approach for the front end to a mobile telephone system. The handsfree module consists of a soft voice switch module (called a weighting balance) and a local comfort noise generator governed by the VAD. On the transmit side, the input signal is pre-processed by down scaling and high pass filtering to remove the long-term DC components. The send input signal level is estimated together with the background noise level. The receive speech level is correspondingly estimated. The activity state then is decided which controls the final attenuation of both send and receive signals such that the local acoustic echo is optimally reduced and likewise the throughput of the genuine send and receive signals is maintained. The comfort noise is added to the send signal to mask any switching artifacts. This is shown in figure-2 which is elucidated in section (2).

An alternative operation of this module is the so called *handset-mode*. This uses the same algorithm as that of the handsfree mode but in a much rudimentary form. It is intended for use in order to reduce the internal acoustic echo within the hand-held mobile telephone instrument. Here the bilateral voice switch is replaced by a single switch in the microphone path only. This introduces a fixed attenuation of nominally 6dB during signal activity from the far-end subscriber. This attenuation value is adapted from its nominal value by the system controller during normal operation. During predominant signal activity from the near-end subscriber, no attenuation is applied in either the receive nor the transmit path. There is also no comfort noise inserted into the transmit path. This is shown diagrammatically in figure-1.

Figure 1: Switching Algorithm in Handset Mode

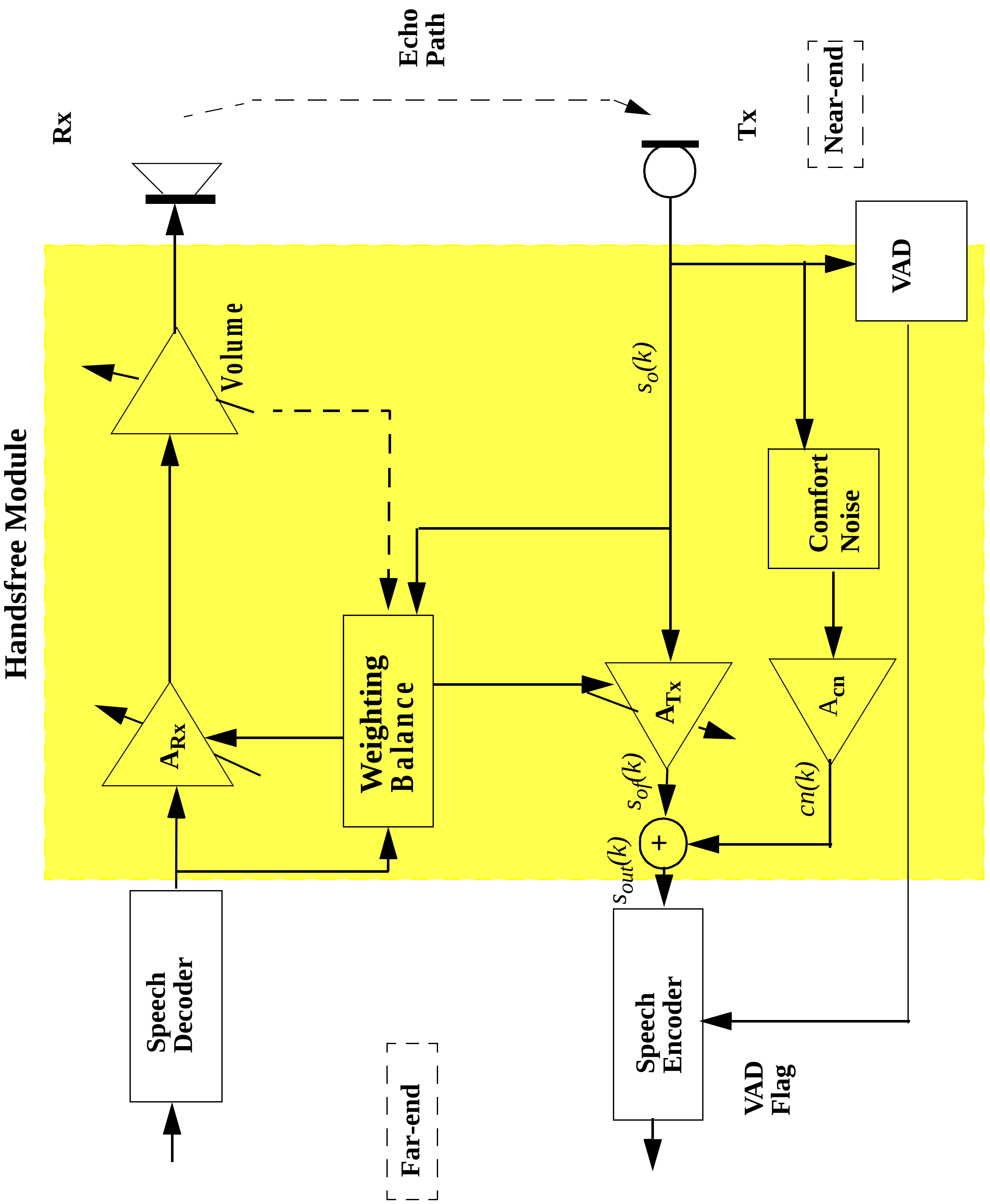


The main differences are highlighted in table 2.

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

Figure 2: . The Handsfree Module



All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

Table 2: Salient Differences between Handfree and Handset operational modes

Handsfree (car kit) mode	*Handset mode*
dynamic acoustic attenuation estimation	fixed acoustic attenuation estimation
use background level estimation for total attenuation calculation	fixed total attenuation
attenuation of microphone signal or loudspeaker signal depending upon the activity detection	attenuation of the microphone signal only if far active is detected
comfort noise insertion	no comfort noise insertion

2.2. The Handsfree Module

In the handsfree module of figure 2, the speech signal from the far end is decoded and passed through the variable attenuator having a loss of a_{Rx}, of between -50dB through to 0dB, see equations (2.12) and (2.13). The resultant signal is fed to the output loudspeaker via the manual volume control. The speech signal from the near-end subscriber is firstly attenuated by a complementary loss of a_{Tx} such that the loop gain never exceeds a predetermined value, see section (2.2.1.2). This signal is passed to the speech encoder and conveyed to the far end subscriber. The attenuation factors are regulated by a weighting balance which apportions the minimum attenuation in the path with overall maximum signal energy. Thus when the far-end subsciber predominately speaks, then his speech will gain the most weighting at the loudspeaker, the converse is true for the near-end subscriber. This is so in order to minimise the return propagation of local echo to the far-end subscriber. This alleviates the unpleasantness of echoed far-end speech and the possibility of instability of the baseband telephone system. In order to provide subjective continuity of the background noise to the far-end subscriber, comfort noise is synthesised and mixed such that the overall energy of the speech and the noise transmitted back is fairly constant (see figure 1). A voice activity detector VAD is added to control the updating of the comfort noise generator and to serve as part of the speech encoder for "DTX ON" purposes.

2.3. Detailed Design Description

The weighting balance consists of the following units which are elucidated here in detail:

2.3.1. Pre-Processing

The input near-end signal *s(k)* in figure-2 is firstly down scaled as

$$s_o(k) = 0{,}5 \cdot s(k) \qquad \textbf{(2.1)}$$

and then offset compensation is applied by a notch filter with the following transfer function

$$H_{of}(z) = \frac{z-1}{z-\alpha} \qquad \textbf{(2.2)}$$

This removes the offset of the input signal $s_o(k)$ to produce an offset free signal $s_{of}(k)$

$$s_{of}(k) = s_o(k) - s_o(k-1) + \alpha \cdot s_{of}(k-1) \qquad \textbf{(2.3)}$$

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

2.3.2. Level Estimation

The average current level *P(n)* in frame, *n*, of 160 samples in the speech frame is estimated according to:

$$P(n) = \frac{1}{160}\sum_{1}^{160}\left|s_{of}(160n+k)\right| \tag{2.4}$$

2.3.3. Logarithm

An approximation of the logarithmic function $20\ log_{10}(x)$ uses a normalised straight-line approximation between the interval 0.5 to 1.0. If $x = P(n)$, then the characteristic can be used such that:

$$log_2(x) = log_2(x_{norm}\cdot 2^n) = n + log_2(x_{norm}) \tag{2.5}$$

Where $x_{norm} = norm(x) = x.\ 2^{-n}$. The actual straight line approximation used is given by

$$(1/100)\cdot 20log_{10}(x_{norm}) = 0{,}2\cdot log_{10}(2)\cdot log_2(x_{norm}) \approx G_a + G_b\cdot x_{norm} \tag{2.6}$$

and the gradient G_b and intersection G_a of the straight line approximation are calculated from the least mean square linear regression formula:

$$min\left(\int_{1/2}^{1}(0{,}2\cdot log_{10}(x_{norm}) - G_a - G_b\cdot x_{norm})^2dx\right) \tag{2.7}$$

The resultant values are: G_a = -0.1174 and G_b = 0.115. The function *LOG(x)* is then given by

$$LOG(x) = 0{,}2\cdot log_{10}(2^n\cdot x_{norm}) = 0{,}2\cdot log_{10}(2^n) + 0{,}2\cdot log_{10}(x_{norm}) \tag{2.8}$$

$$= 0{,}2\cdot n\cdot log_{10}(2) + G_a + G_b\cdot x_{norm} \tag{2.9}$$

2.3.4. Line_act_near & Line_act_far Functions

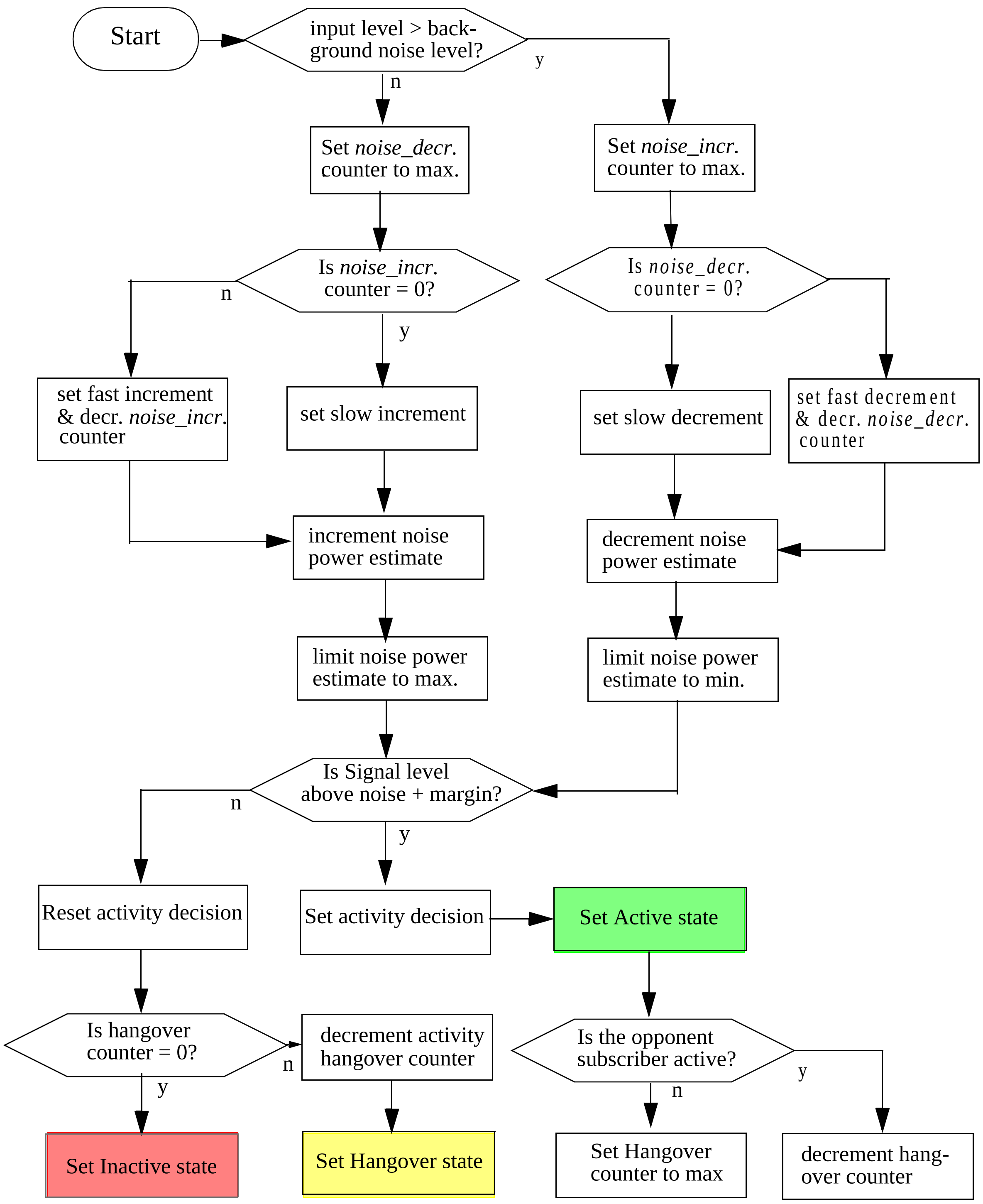
These functions make an attempt to decide from where the speech activity occurs. It is important here to produce an accurate estimation of the background noise level. Speech activity is then deemed to be present when the current level exceeds the background noise level plus a specific margin offset. The estimated background noise level slowly increases with large current signal amplitudes and slowly decreases when the input signal level reduces. Counters are employed to apply hangover with specific step-sizes, see table-3.

Figure 4 shows an excursion of the input signal power and associated estimated noise and switching thresholds in a diagrammatically explanatory form. The lower line represents the estimated acoustic background noise level and the upward displaced line is the same with a constant margin added to it. When the input signal power rises above the current noise level then the estimate of the noise power slowly increases together with the switching threshold and speech activity is deemed to exist. Conversely, when the input signal level falls below the current switching threshold, then the noise power estimate and the switching threshold also decrease at a rate larger than the incremental gradient. Subsequent to each active phase, a fixed hangover period is augmented. The decision then goes into an inactive state until the end of the current speech pause. The flow logic is shown in figure 3. It can be seen that in the speech activity phase, the background noise

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

Figure 3: Decision Flow Chart of LINE_ACTIVE_NEAR/FAR Process

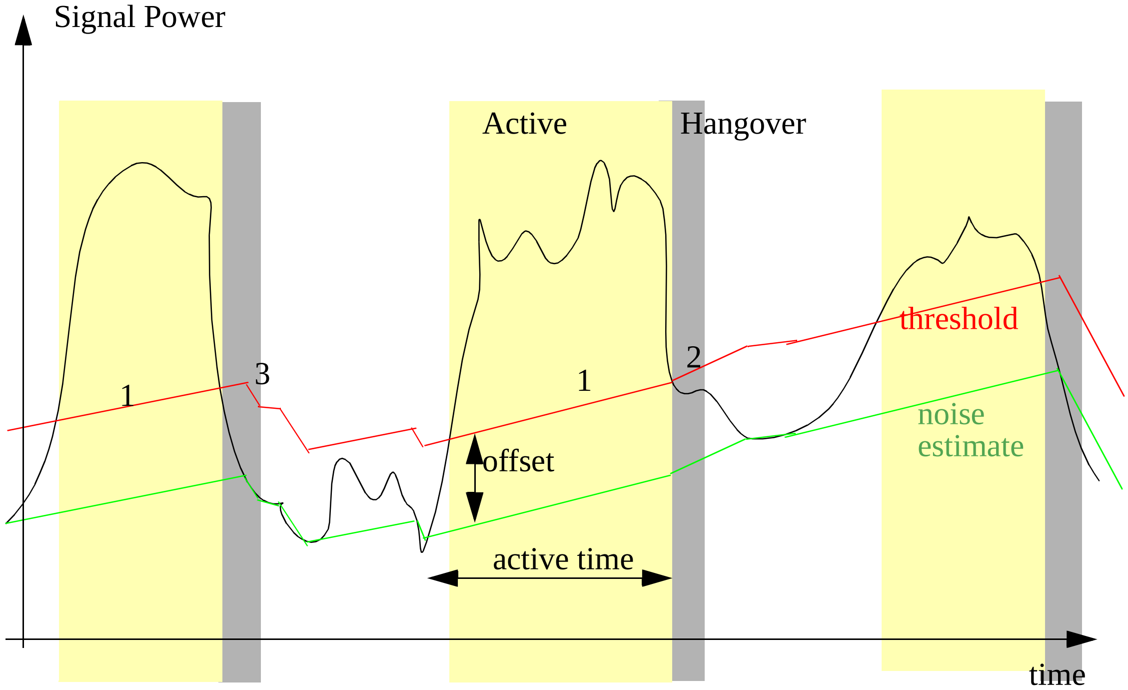


All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

increment counter changes from slow to fast increment when the increment counter attains the value *MAX_CO_HAOV*. This is to accommodate more rapid changes in background noise levels.

Figure 4: Temporal Signal Level Variations.



1 = slow increment, 2 = fast increment and 3 = slow decrement

As can be seen from figure 4, the actual activity threshold level is time dependant which in turn depends upon the respective instantaneous noise counter value. The noise threshold is initialised to *MIN_LEV_BGR*. This value is input via the system controller and can be programmed by the product provider based upon field tests. The value used for the simulation is shown in table 3.

2.3.5. Activity Determination

This decides which overall state currently ought to be appropriate (i.e. near active, far active, double-talk or idle). This occurs simply by comparing signal activity. When both signals are active, it must be determined if the double-talk situation is apparent or if the near-end signal is just the echo from the far-end signal by comparing the signal level with the estimated echo level. The decision flow chart to determine the respective half duplex state is shown in figure 4. The (coloured) shaded boxes indicate the final decision states.

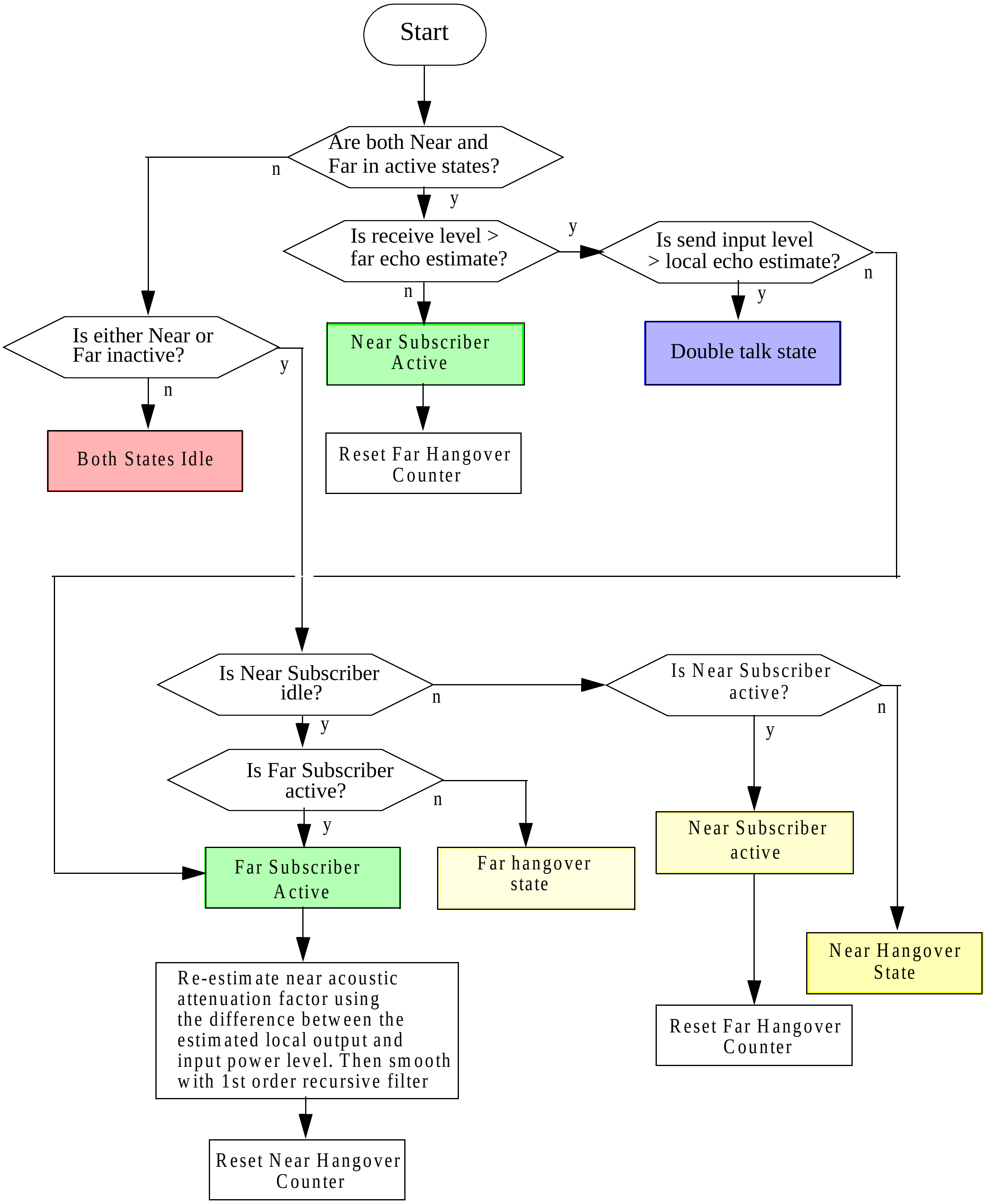
ALPHA_AC is the smoothing constant used for the estimation of the acoustic attenuation value *ATT_AC*. This is only used for the handsfree (car-kit) mode. The acoustic attenuation value is calculated by:

$$ATT\_AC = (1.0 - ALPHA\_AC) * ATT\_AC\_OLD + ALPHA\_AC * ALPHA\_AC\_MEAS \quad \textbf{(2.10)}$$

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

Figure 5: Decision Flow Chart of ACTIVE routine



All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

2.3.6. Weight Function

This function is only used for the *handfree* mode as in section (2.2). It calculates the appropriate weighting functions (attenuation levels, $A_{Tx}(n)$ and $A_{Rx}(n)$) applied to the weighting-balance. $A(n)$ is in fact a function of the estimates of the levels of the background noise and the near-end acoustic echo. $A(n)$ is constant for the whole of the 20ms-speech frame, n. The switching depth att_{tot} in dB is calculated as

$$att_{tot} = att_{corr} + echo_{est} - noise_{est} \tag{2.11}$$

where att_{corr} = attenuation correction value in dB

$echo_{est}$ = current estimated local echo attenuation

$noise_{est}$ = current estimated background noise level

The currently active subscriber always immediately receives the minimum attenuation (i.e., 0dB) and the inactive signal is muted to att_{tot} dB. The immediacy prevents voice clipping at the onset of speech bursts. For the idle or double talk states, both paths receive an attenuation of $att_{tot}/2$ dB which is the quiescent status of the weighting balance. In the idle state, the attenuation in both paths is varied accordingly by 2 dB per 120 ms which is controlled by the IDLE-counter.

If in the receive path, the attenuation a_{Rx} is applied, then $a_{Tx} = att_{tot} - a_{Rx}$ is applied into the transmit path such that the near-loop attenuation is maintained at att_{tot}. The value of att_{corr} is loaded via the system controller and remains constant until the next reset. Its value is programmed by the product provider and can be based upon field tests. The value of 6 dB is used for the simulation.

2.3.7. Normal Weighting Function

This is a variant of the weighting function which is used exclusively for the *handset* mode. Here the microphone input speech is attenuated initially by a fixed amount of 6 dB when the far-end subscriber is active. This fixed attenuation is subsequently varied by the system controller during normal operation. The loudspeaker output level is not adjusted. The local attenuation estimate, att_{tot}, is also fixed. The meaning of att_{tot} depends upon the operation mode (handset or handsfree). In handsfree mode, this is represented as the value/100 dB. For example, when att_{tot} needs to be decreased by 10 dB, the value -0.1 needs to be programmed. In handset mode, the programmed value is represented by att_{tot} divided by 2. For example, when a total attenuation of 10 dB is requested, the programmed value should be 5.

2.3.8. Path attenuation

In handsfree mode, the corresponding attenuation factor is accordingly applied in each path on a sample by sample basis, viz:

$$A_{Tx} = 10^{(a_{Tx}/20)} \tag{2.12}$$

and

$$A_{Rx} = 10^{(a_{Rx}/20)} \tag{2.13}$$

Thus:

$$s_{out}(160n + k) = A_{Tx} \cdot s_{in}(160n + k) \tag{2.14}$$

and

$$r_{out}(160n + k) = A_{Rx} \cdot r_{in}(160n + k) \tag{2.15}$$

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

The weighting balance is manifested in such a way that the values of A_{Tx} and A_{Rx} are pointers to a 26-entry gain table, *att[.]*, spanning from 0 to 50 dB in 2-dB steps.

Table 3: Constants used for the Handsfree Weighting Balance

Constant Name	*Description*	*Nominal Value*
THRESHOLD	Margin of Input level above Noise floor	9 dB
ATT_CORRECTION	see section 2.3.6	6 dB
CN_FACTOR	Output amplification factor to mixer for Comfort Noise Generator	0.5
TOTAL ATTENUATION	Initial value for the fixed attenuation estimate between loudspeaker and microphone	6dB, updated by the system controller
INC_CO_HAOV	Step size for incremental hangover counter for changing from slow to fast incr.	4 (= 80ms)
DEC_CO_HAOV	Step size for decremental hangover counter for changing from slow to fast decr.	1 (= 20ms)
FAST_INCREMENT_BGR	incremental step	37
FAST_DECREMENT_BGR	decremental step	147
SLOW_INCREMENT_BGR	incremental step	12
SLOW_DECREMENT_BGR	decremental step	49
WEIGHT_STEPS	Step for attenuation correction	2 dB
ALPHA_AC	Smoothing Constant for background noise estimate	0.0066
ALPHA_AC_1	Smoothing Constant for same	0.993
FAST_ALPHA_AC	Smoothing Constant for same	0.0066
FAST_ALPHA_AC_1	Smoothing Constant for same	0.993
MAX_AT_AC	maximum attenuation	15dB
MAX_CO_HAOV	max counter to change from slow -> fast incr./decr.	8 (Near and far end)
MAX_LEV_BGR	maximum defined background noise level	60dB
MIN_LEV_BGR	minimum define background noise level	40dB

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

Table 3: Constants used for the Handsfree Weighting Balance

Constant Name	*Description*	*Nominal Value*
CO_IDLE_MAX	max value before	10
CO_DEC_HO, CO_INC_HO	hangover count	10 (Near and far end)
STATE_0	idle	0
STATE_1	near hangover	Debugging Labels Only
STATE_2 & STATE_3	near active	
STATE_4	far hangover	
STATE_5 & STATE_6	far active	
STATE_7 & STATE_8	double talk	

2.3.9. Loudspeaker Volume Information

The audio output volume of the mobile telephone handset is manually controlled via the microcontoller on the hardware. For the C-reference simulation, an appropriate file contains values pertaining to the volume control. The setting of this directly influences the estimated echo attenuation, $echo_{est}$. and informs the handsefree module accordingly. The corresponding simulated values of the volume attenuation factor are incorporated in the file *LOUD.ILS*. This file contains values of k = 0 to 7 (3 LSBs) which are used as indices for the table *loudness_table[k]* in program file *sp_hf.c*. The values begin at 0 dB and increase corresponding to 3 dB steps of attenuation. This is added to $echo_{est}$ in each 20 ms-speech frame and varies only when k is effectively manually altered (i.e. if k is held constant from one frame to the next, then no volume correction change occurs). The 9th bit is ordered according to the following operational modes

Table 4: Operation Mode

	Handsfree Mode (Car Kit)	*Handset Mode*
Bit 9 in file LOUD.ILS	0	1

2.3.10. Comfort Noise Generation

A block schematic diagram of the comfort noise generator is shown in figure 6. The input speech signal is taken from the near-end subscriber. Initially, the first 9 autocorrelation coefficients are computed. These are used to produce the reflection coefficients used later for the adaptive filter. The comfort noise is synthesised by a random noise generator having a flat spectrum which in turn is shaped by the adaptive filter. The spectral shaping is applied continuously to the input white noise however, the coefficients are only updated when the VAD switches to the non-speech decision. This is intended such that the comfort noise simulates the spectrum of the input audio noise to the near-end microphone. The coefficients are interpolated by a first order recursive filter in order to smooth out any abrupt steps that might occur when the updating is interrupted by the VAD. Similarly, the gain factor which is obtained by square rooting the zero lag autocorrelation value, is also interpolated to ensure smooth transitions. The de-emphasiser is required to model the microphone pickup of the near-end audio noise.

Copyright PHILIPS GmbH / TCMC 1999 PHILIPS INTERNAL ABD08-G27-DD013

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

Figure 6:

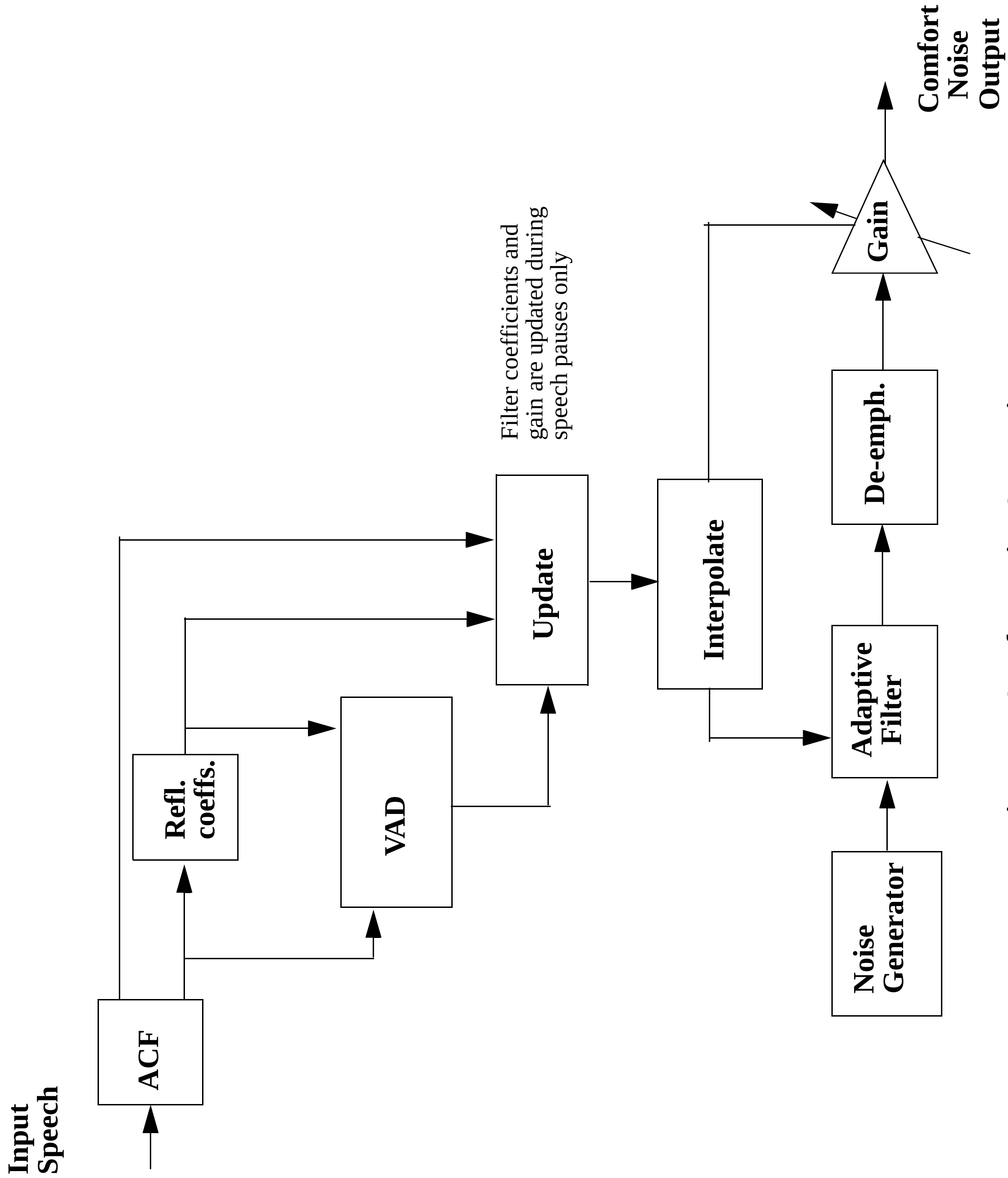


Figure-5: Comfort Noise Generation

All rights reserved. Reproduction in whole or in part is
prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur
zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

2.3.10.1. Pre-emphasis

To preprocess the input speech samples, s(k), these are fed through a first order filter, viz:

$$s_o(k) = s(k) - a \cdot s_o(k-1) \quad \textbf{(2.16)}$$

where a = 0.86

2.3.10.2. Parcor coefficients

The first 9 values of the autocorrelation function are calculated according to equation (3.2) in ref.[4], subsequently, the Parcor-LPC-coefficients *r(m)*, where *m*=0 to 8, are computed using the Schur Recursion algorithm, as depicted in the flow chart of figure 3.2 of ref.[4]. These then are smoothed by the interpolation filter viz:

$$r_{cn}(m, n) = 0{,}9 r_{cn}(m, n-1) + 0{,}1 r_{cn}(m, n) \quad \textbf{(2.17)}$$

where *m* is the respective smoothed reflection coefficient number

and *n* refers to the current speech frame number.

2.3.10.3. Gain Computation

The energy of the input signal is updated only when no speech is present in the input signal as determined by the VAD-module. Thus an estimate of the background noise power is made. Firstly, the transversal LPC coefficients are recalculated from the reflection coefficients by the *Step_up* procedure as detailed in section (3.3.2), page 20 of ref.[5]. The autocorrelated predictor coefficients are computed int the routine *compute_rav1* in section 3.3.3, page 20 of ref.[5]. The energy of the LPC-filtered signal is estimated as the convolution of the autocorrelation function of the input signal $\varphi_{ss}(k)$ and the autocorrelation of the impulse response of the LPC filter $\varphi_{hh}(k)$.

The filtered signal energy is then given by:

$$E_y = \varphi_{hh}(0) \cdot \varphi_{ss}(0) + 2 \cdot \sum_{m=1}^{8} \varphi_{hh}(m) \cdot \varphi_{ss}(m) \quad \textbf{(2.18)}$$

The values of *E_prod* and *M_prod*, as obtained by the resultant from the C-routine *energy_computation*, represent the energy estimate of the filtered speech frame which therefore must be square-rooted to give the gain factor, m_{fac}. This is performed by halving the exponent E_prod and applying an approximated 2nd-order 16-bit square-root function to the mantissa *M_prod*. If *E_prod* is odd, then one is added to this prior to square-rooting and consequently *M_prod* is multiplied by $\sqrt{0{,}5}$ such that the result is always in standard form.

The resultant *mfac* is de-normalised and smoothed according to the interpolation formula:

$$gain(n) = 0{,}9 m_{fac}(n-1) + 0{,}1 m_{fac}(n) \quad \textbf{(2.19)}$$

where *n* refers to the current speech frame

2.3.10.4. Noise Generation

This is produced using a pseudo-random number sequence from a 31-bit feedback tapped shift-register. The resultant output number series has the range of 0x8000 to 0x7fff. The generator pol-

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

ynomial has the form:

$$y = 1 + x^{28} + x^{31} \quad \textbf{(2.20)}$$

The shift register is initialised with an arbitrary value for the lower 16 bits with 12345_{10} and with 6789_{10} for the upper 16 bits. The output samples from the noise generator are multiplied by the gain value, *gain(n)*, before being fed to the synthesis filter.

2.3.10.5. Synthesis Filter

The flat spectrum random noise is shaped by a corresponding lattice filter, as shown in figure 3.3 and explained in section (5.3.4) of ref.[4], using the smoothed reflection coefficients, $r_{cn}(m,n)$. This is performed in order to represent the audio ambient noise spectrum at the near subscriber's terminal.

2.3.10.6. De-emphasis and Upscaling

The samples are de-emphasised according to:

$$cn(k) = d(k) + 0{,}86 \cdot cn(k-1) \quad \textbf{(2.21)}$$

The *cn(k)* samples are then used as the synthesised comfort noise for frame *n* and correspondingly mixed into the signal, $s_{out}(k)$, from the near subscriber's terminal. Equation (2.15) is therefore modified according to:

$$s_{out}(k) = 2{,}0 \cdot [A_{cn} \cdot cn(k) + A_{Tx} \cdot s_{in}(k)] \quad \textbf{(2.22)}$$

In order to keep the energy of the transmitted signal roughly independent of the value of A_{Tx}, the following is used:

$$A_{cn} = 1{,}0 - A_{Tx} \quad \textbf{(2.23)}$$

2.3.11. VAD

The operation of the Voice-Activity Detector (VAD) is copiously described in ref.[5] to which the reader is referred. This form of voice-switching is used in addition to the activity detectors of the handsfree algorithm, firstly to attain an even more reliable operation and secondly, the VAD-module is already present in the GSM-speech coder and is therefore provided with no extra firmware code capacity requirements. The VAD employs the autocorrelation values and the parcor coefficients to deliver the VAD-flag for the current speech frame. The pitch value, *ptch*, is set to zero for handsfree operation. When the mobile telephone terminal is in handsfree mode, the VAD-decision flag is propagated to the speech encoder which is used subsequently for DTX ON purposes, as shown in figure-7.

3. Software Description

This general description of the software pertains only to the C-reference code of which the main files and their respective routines are listed in table-5. The operational description of the assembler code is briefly indicated in the interface document ref.[3] and is not elaborated here.

The associated file-names, which are fixed for the sake of convenience, are shown for the overall simulation system in figure 7 and listed in table 5. The routine calling trees of the C-reference code of the handsfree module and the comfort noise generation are shown is figure 8 and 9. There are

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

of course lower level mathematical subroutines used, which are directly taken from the GSM full-rate speech coder described in ref.[4]. The reader is referred to the appropriate program listings of the 'SP'-Modules to enable insight into the operation of these.

Table 5: Source Files and associated Routines

File	*Description*	*Routine(s) contained therein*
HF.C	Main Calling Program	
SP_HF.C	Soft Voice Switch (Weighting Balance)	encoder_handsfree decoder_handsfree active av_abs_val logarith line_act_near line_act_far calc_lsout calc_lrout hf_reset
FRCN_GEN.C	Comfort Noise Generation using subroutines from the GSM full-rate speech codec	fr_cn_gen par_mean CN_pre_emp wurzel CN_ran_asl CN_syn_ltc CN_de_emph
VAD.C	Voice Activity Detector	energy computation average_acf schur_recursion step_up compute_rav1 predictor_values spectral_comparison tone_generation threshold_adaptation vad_decision periodicity_update vad_algorithm

Table 6: I/O and Test Files For the Handsfree Simulation

Programmed File Name	*Description*
s_in.ils	Sampled Input Speech from Near Speaker
s_echo.ils	Sampled Near Acoustic Echo Signal

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

Table 6: I/O and Test Files For the Handsfree Simulation

Programmed File Name	***Description***	
s_ref.ils	= *s_in.ils* + *s_echo.ils* input to Handsfree Unit	
s_out.ils	Output Sampled Signal fro Handsfree Unit	
sp_cod.ils	Coded Speech Signal from Speech Encoder (Not part of the Handsfree Unit)	
sp_dec.ils	Coded Speech Signal to Speech Decoder (Not part of the Handsfree Unit)	
r_in.ils	Input Sampled Speech from Far Speaker to Handsfree Unit	
r_out.ils	Output sampled Speech Signal from Handsfree Unit To Near Subscriber	
i_hf_fr.ils	Output Debug Test Files to Test Assembler Firmware. For Format, see Test Specification	
o_hf_fr.ils		
loud.ils	Volume Setting per 20ms Speech Frame (0 - 7)	Bits 0 -2
	Operation Mode: 0 = Handsfree, 1 = Handset	Bit 9

3.1. Data Format

All sampled I/O-speech is stored in 16-bit form on ILS-format files. The samples of the files *r-in.ils* and *s_ref.ils* at the input to the handsfree-unit are masked by 13 left-hand justified bits. The 3 LSBs are set to zero. In the output file *s_out.ils*, the multiplexed VAD- and handsfree-active flags in bits 0 and 1, respectively.

3.2. Test Sequences

The test sequences of speech signals used to test the handsfree simulation are listed in table 7

Table 7: Test Sequences

Data File Name	***Description***
frau.ils	female clean speech (German)
kind.ils	child's clean speech (German)
mann.ils	male clean speech (German)
no_speech-ils	non speech, no acoustic noise
vad_fem.ils	female noisy speech with pauses (English)
vad_mal.ils	male noisy speech with pauses (English)

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

Figure 7:

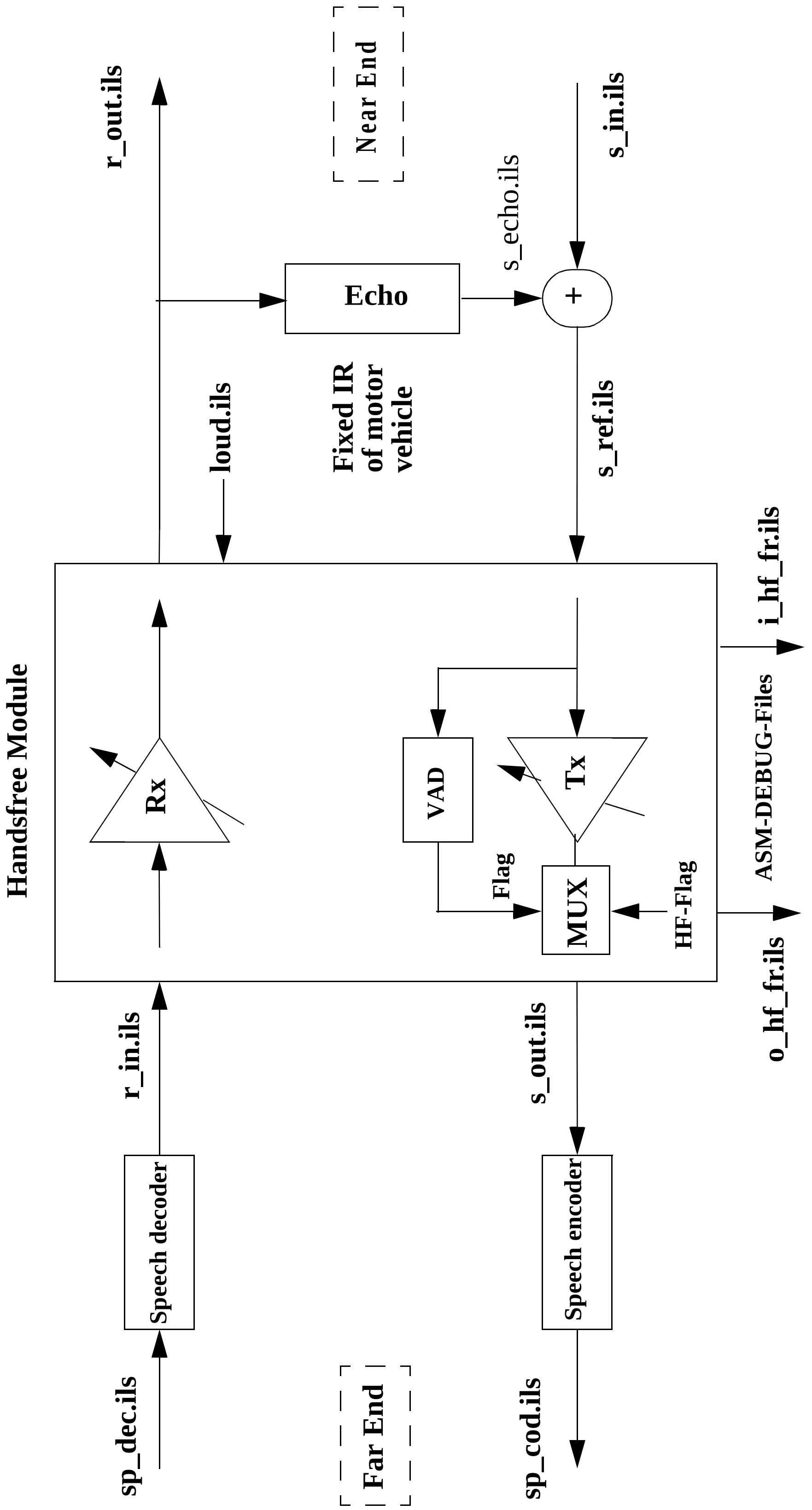


Figure-7: I/O-File Handling

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

Figure 8:

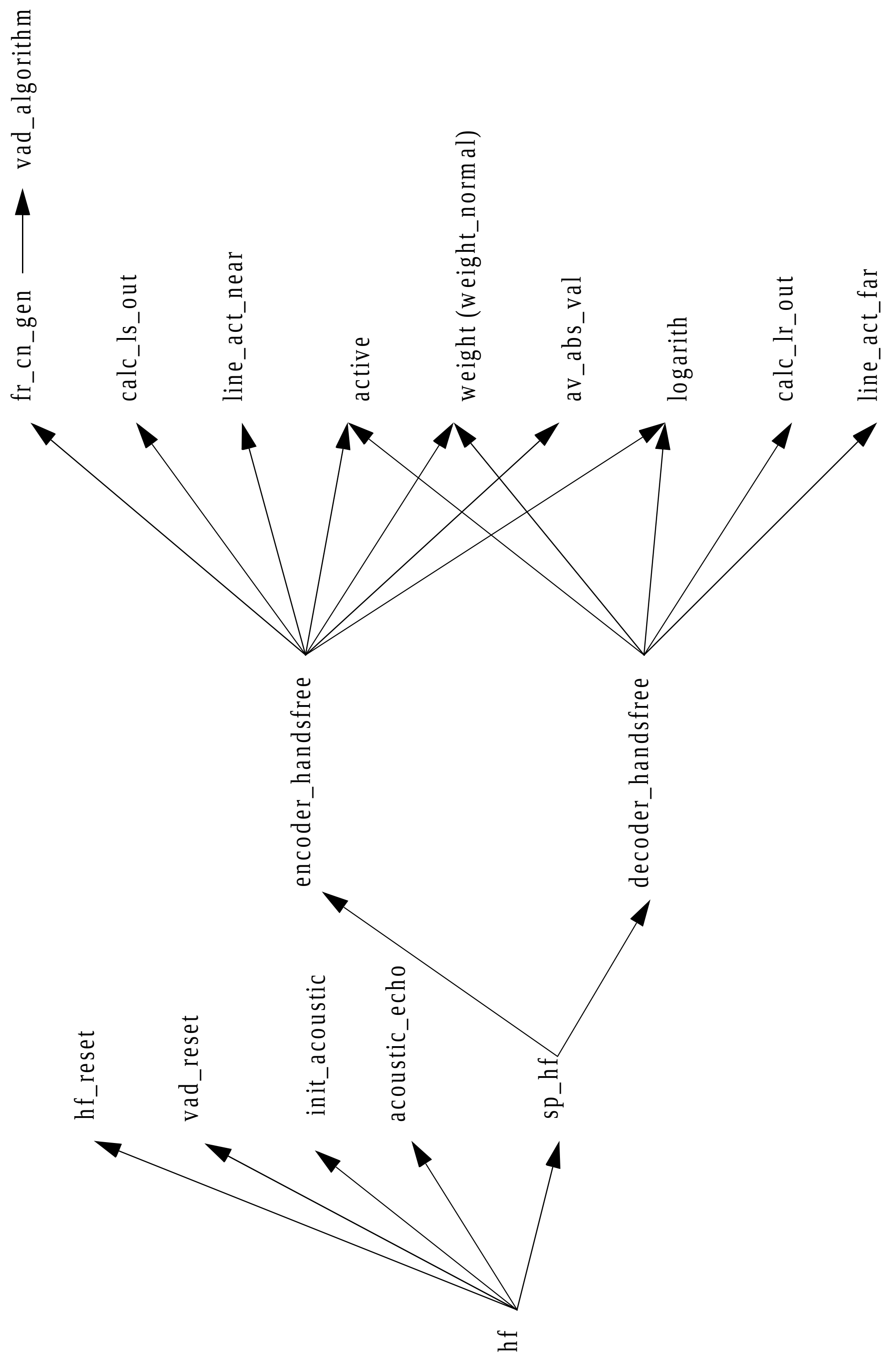


Figure-8: Calling Tree for the Handsfree Module

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers

Figure 9:

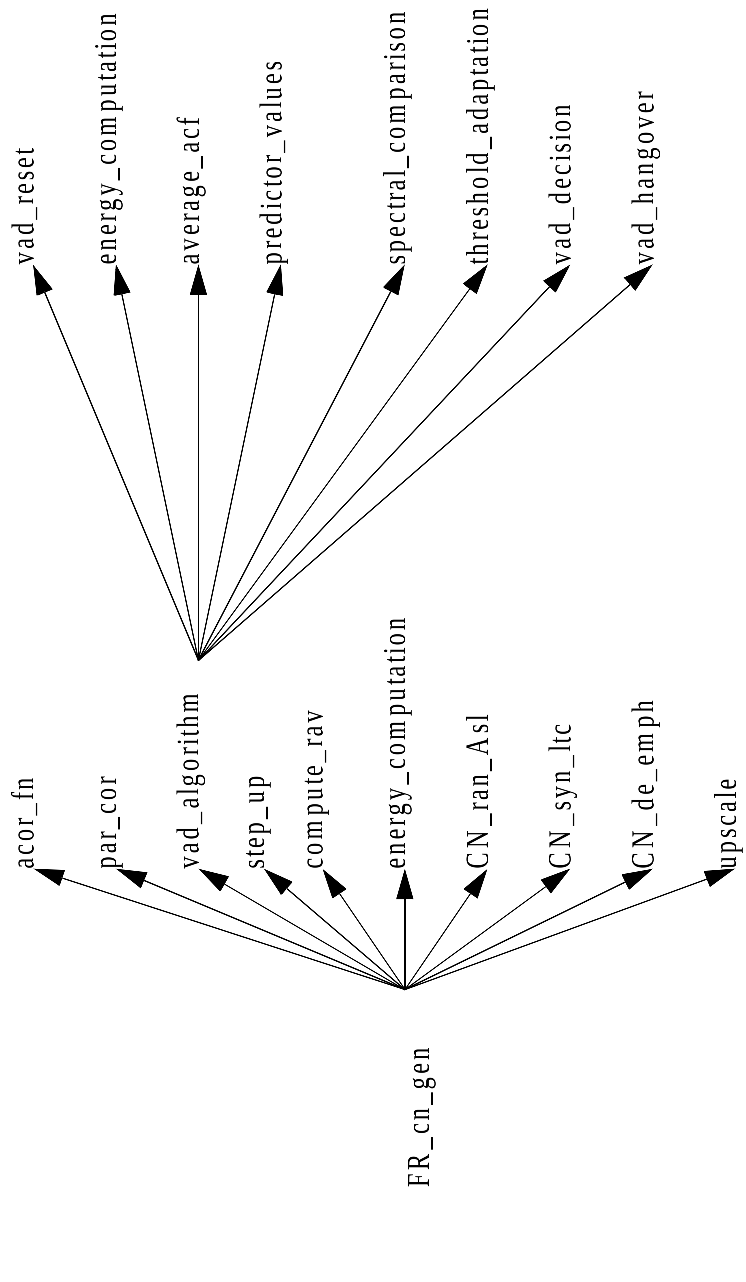


Figure-9: Comfort Noise Generation Controlled by the VAD

All rights reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner

Alle Rechte vorbehalten. Vervielfältigungen ganz oder teilweise nur zulässig mit schriftlicher Zustimmung des Urheberrechtsinhabers