AN ALGORITHM FOR CANCELLATION OF SIDETONE OSCILLATIONS

Martin Schönle, Virginie Gilg

Siemens AG, ICM Mobile Phones, Grillparzerstrasse 10-18, 81675 Munich, Germany
martin.schoenle@siemens.com

ABSTRACT

An algorithm for detection and cancellation of sidetone oscillations in mobile phones is presented. The sidetone signal is represented by a 4^{th} order LPC polynomial, whose root structure is analysed by a computationally efficient two-stage approach. In difference to LMS based adaptive filters this approach offers a fast and reliable detection of evolving oscillations. A notch filter with small bandwidth is used for cancellation and therefore almost no distortion is introduced to the speech signal.

1. INTRODUCTION

Design of the electro-acoustic parameters defining the sidetone path in modern mobiles phones has become a critical issue during the last years. Reasons are decreasing enclosure sizes, low-cost acoustic and mechanical components, and the usage of acoustic resonators to increase earpiece performance. As it is shown in Fig. 1, an electro-acoustic feedback loop may evolve. One part of this loop is intended and defined by the sidetone path: The microphone signal, after some filtering and attenuation, is fed back to the loudspeaker path to provide the user with a convenient perception of his own voice.

The feedback loop may be closed by mechanical coupling between earpiece and microphone via the enclosure, mainly if these devices are mounted on the same side of the enclosure. In addition acoustic coupling may be caused occasionally by some users, who simultaneously cover both devices with their hand. This can result in annoying whistling, even during a phone call and - in the extreme case - damage the hearing capability of the user.

The generated resonances are dependent on the properties of the feedback loop and the resonance frequencies are varying in time, therefore an adaptive algorithm has to be used. One challenge is the detection of evolving oscillations during the presence of a concurrent speech signal. The other challenge is to provide a fast algorithm, which is able to detect and cancel oscillations before they are audible. In addition, the detected oscillations must be cancelled with only small distortion of the speech signal.

Adaptive cancellation of feedback signals is also used in the domain of hearing aids. It is known that wideband adaptive algorithms like the LMS cause distortions because input and output signals of the adaptive filter are mainly speech signals and therefore strongly correlated, e.g. [1], [2]. Distortions typically occur at low frequencies, where most of the speech energy is concentrated. In the same manner LMS-controlled adaptive notch filters tend to concentrate their notch frequencies at regions of high energy. For speech these are the formant regions, whereas oscillations with higher frequencies might not be detected. A drawback of all LMS-style algorithms is their slow convergence.

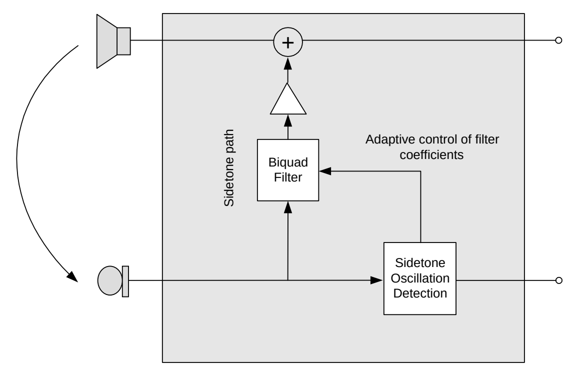

Fig. 1. Digital sidetone path in mobile phones

Therefore a different approach is proposed in this paper. It is based on a 4^{th} order LPC analysis of the sidetone signal which is able to separate speech poles (formant frequencies) and oscillation poles with resonance frequencies above 2 kHz. A two-stage algorithm is chosen for the analysis of the root structure of the LPC polynomial. In a first step the LSF (Line Spectral Frequency) representation of the LPC polynomial is computed and evaluated [3, 4]. In a second step the pole frequency of critical poles (possibility of oscillation) is used as parameter to compute a start polynomial for the Bairstow Algorithm, which provides a robust and efficient method for the computation of the LPC roots [5, 6]. The position of these roots in the z-plane is used for the detection and cancellation of oscillations. If they appear in the sector marked as critical in Fig. 2, the pole frequency of these roots is used as center frequency of an adaptive notch filter, which can be driven into the signal path and moved out again. As it is shown in Fig. 1 the adaptive notch filter is implemented as a standard 2^{nd} order biquad filter. The notch filter transfer function is used in a representation, which allows a description by the parameters pole frequency, gain, and 3dB-bandwidth [7]. The coefficients of this filter are controlled by the analysis and detection algorithm, which is introduced in the next chapter.

2. ALGORITHM DESCRIPTION

2.1. Basic detection principle

The algorithm is based on a 4^{th} order LPC analysis of the sidetone signal. The short-term sythesis filter is given by

$$S(z) = \frac{1}{A(z)} = \frac{1}{\sum_{i=0}^{4} a_i z^{4-i}} \quad . \tag{1}$$

The 4^{th} order polynomial $A(z)$ has real coefficients and can be factorized into 4 root factors. The roots are either real roots or conjugate complex root pairs. Only the complex roots are of interest as the relevant frequencies are situated around 1 kHz (speech formant frequencies) or above 2 kHz (oscillation frequencies). By computing the pole distribution of the polynomial $S(z)$ a statement about the character of the observed signal can be made. If only speech components are present, one or both of the conjugate complex pole pairs are moving to the formant frequencies. If an oscillation is evolving, one pole pair moves to the oscillation frequency even during speech sequences, because a sine oscillation can exactly be modelled by one conjugate complex pole pair. The other pair stays at the formant frequencies. Thus speech and oscillation can be separated by looking at the location of the LPC poles in the z-plane.

2.2. LPC Analysis

The LPC polynomial $A(z)$ is computed by the Levinson-Durbin Recursion as it is described in the GSM standards for the EFR or AMR speech codecs, e.g [4]. The autocorrelation coefficients required for the Levinson-Durbin Recursion are computed over a block of samples. To obtain smoothed values, the autocorrelation coefficients of the previous block are considered recursively:

$$\begin{aligned} R_{xx}(i)_{new} &= \sum_k x(k)x(k+i),\ i = 0\ldots 4 \qquad (2) \\ R_{xx}(i) &= \alpha R_{xx}(i) + (1-\alpha) R_{xx}(i)_{new} \\ k &= 0\ldots N-1, \end{aligned}$$

where N is the block length and α the recursion coefficient.

2.3. Computation of LPC roots

There are several methods to compute the roots of a 4^{th} order polynomial exactly. As most of them use computationally complex matrix inversions or eigenvalue approaches neither is suited for a real-time implementation. Therefore a two-step approach is proposed: First of all the LPC representation is transformed to a LSF representation [3, 4]. The results of the LSF polynomial evaluation are used as input values for the Bairstow Algorithm [5, 6], which is a robust and efficient method to find the quadratic factors for the complex conjugate roots of a polynomial with real-valued coefficients. As we are interested in one complex conjugate pole pair only, which additionally has to be situated in a well-defined sector of the z-plane, we can pre-select the output of the LSF root finding algorithm, compute a 2^{nd} order polynomial corresponding to a critical LSF root pair, and use this polynomial as an input to the Bairstow Algorithm. The main drawback of the Bairstow Algorithm, the difficulty to find good initial approximations [6], is circumvented by this approach.

2.3.1. *Line Spectral Frequencies*

A symmetric polynomial $F_1(z)$ and an antisymmetric polynomial $F_2(z)$ are formed from the LPC polynomial $A(z)$ by adding and subtracting the time-reversed system function:

$$\begin{aligned} F_1(z) &= \frac{A(z) + z^{-5}A(z^{-1})}{1+z^{-1}} \\ F_2(z) &= \frac{A(z) - z^{-5}A(z^{-1})}{1-z^{-1}} \end{aligned} \tag{3}$$

The polynomials have fix zeros at $z = -1$ and $z = 1$, respectively. As they are not of interest, they are removed by polynomial division, as it is shown in (3). This results in

$$\begin{aligned} F_1(z) &= z^{-4} + f_{11}z^{-3} + f_{12}z^{-2} + f_{13}z^{-1} + 1 \\ F_2(z) &= z^{-4} + f_{21}z^{-3} + f_{22}z^{-2} + f_{23}z^{-1} + 1\ , \end{aligned}$$

with

$$\begin{aligned} f_{11} &= a_1 + a_4 - 1 \\ f_{12} &= -a_1 + a_2 + a_3 - a_4 + 1 \\ f_{13} &= f_{11} \\ f_{21} &= a_1 - a_4 + 1 \\ f_{22} &= a_1 + a_2 - a_3 - a_4 + 1 \\ f_{23} &= f_{21} \end{aligned} \tag{4}$$

by comparison of coefficients.

Each polynomial is now evaluated on the unit circle. By extracting the linear-phase factor $2e^{-j2\omega}$ and with the mapping $x = cosk\omega$ each polynomial can be written as Chebyshev expansion [3]:

$$F_i(e^{j\omega}) = 2e^{-j2\omega} \sum_{k=0}^{2} \lambda_{ik} cosk\omega \tag{5}$$
$$i = 1, 2 \; ,$$

resulting in the Chebyshev polynomials

$$C_{F_1}(x) = \frac{1}{2} f_{12} + f_{11}x + 2x^2 - 1 \tag{6}$$
$$C_{F_2}(x) = \frac{1}{2} f_{22} + f_{21}x + 2x^2 - 1.$$

Each of these polynomials has 2 zeros on the upper half of the unit circle and it can be shown that they alternate each other, see Fig. 2. The polynomials are evaluated using the algorithm in [3], which computes the results of (6), alternating for $C_{F_1}(x)$ and for $C_{F_2}(x)$ at L equidistant points of the unit circle ($\omega = 0, \frac{\pi}{L}, \frac{2\pi}{L}, \ldots, \frac{(L-1)\pi}{L}$), and looks for sign changes. The final zero position is calculated by interpolation. Four root pairs on the unit circle are the result of the LSF algorithm. It is known [3] that a pair of LSF roots tends to enclose the corresponding LPC roots. It can be seen from Fig. 2 that this is true for the two oscillation poles above the frequency threshold of 2.5 kHz (8 kHz sampling frequency assumed), but also for the two 'formant poles' at low frequencies.

The phase angles of the two LSF roots above the frequency threshold shall be named $\phi_{LSF,1}$ and $\phi_{LSF,2}$, and the start value for the LPC root evaluation is written as $z^s_{LPC} = r^s_{LPC} e^{j\phi^s_{LPC}}$. Only the phase angles of the LSF roots in the upper half of the unit circle are considered for the computation of ϕ^s_{LPC}.

As start polynomial for the Bairstow Algorithm the quadratic factor $r(z) = (z - e^{j\phi})(z - e^{-j\phi}) = z^2 - 2zcos\phi + 1$ is computed with

$$\phi = \phi^s_{LPC} = \frac{\phi_{LSF,1} + \phi_{LSF,2}}{2} \quad . \tag{7}$$

Thus the mean value of the two LSF phase angles is taken (dotted line in Fig. 2). In principle every value $0 < r^s_{LPC} \leq 1$ could be considered as start value for r^s_{LPC}. As we want to detect sine oscillations the best choice is $r^s_{LPC} = 1$.

2.3.2. *Bairstow Algorithm*

Given the polynomial $A(z) = \sum_{i=0}^{4} a_i z^{4-i}$, and the quadratic factor $r(z) = z^2 - pz - q$, with $p = 2cos\phi$ and $q = -1$ from the LSF evaluation, we can write [6]:

$$A(z) = (z^2 - pz - q)(b_0 z^2 + b_1 z + b_2) + \underbrace{b_3(z-p) + b_4}_{remainder} \tag{8}$$

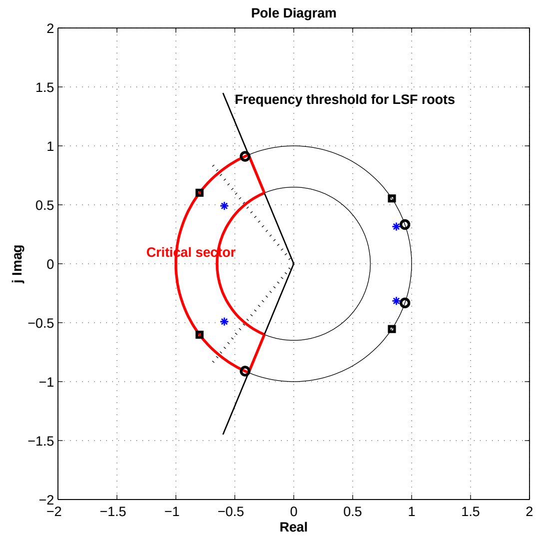

Fig. 2. Basic detection principle: Evaluation of LPC root structure using the analogy between LPC roots (stars in the z-plane) and LSF zeros (diamonds and circles on the unit circle)

The polynomial is represented as multiplication of the known quadratic start factor, an unknown quadratic factor, and a remainder term. This procedure is called synthetic division. By multiplying the quadratic factors on the right side of (8) and comparing the coefficients to the polynomial on the left side we obtain:

$$\begin{aligned} b_0 &= a_0 \\ b_1 &= a_1 + pb_0 \\ b_2 &= a_2 + pb_1 + qb_0 \\ b_3 &= a_3 + pb_2 + qb_1 \\ b_4 &= a_4 + pb_3 + qb_2, \end{aligned} \tag{9}$$

or, recursively,

$$\begin{aligned} b_i &= a_i + pb_{i-1} + qb_{i-2} \\ i &= 0 \ldots 4, \;\; b_{-2} = b_{-1} = 0. \end{aligned} \tag{10}$$

It can be shown, that if $b_0 z^2 + b_1 z + b_2$ is to be another quadratic factor, b_3 and b_4 have to be equal to zero. Both coefficients are non-linear functions of p and q:

$$\begin{aligned} b_3(p, q) &= 0 \\ b_4(p, q) &= 0 \end{aligned} \tag{11}$$

The application of the Newton-Raphson procedure in this context to solve the system of non-linear equations (11) is called Bairstow Method [6]. Applying the partial derivatives with respect to p and q and writing b_3 and b_4 as a 1^{st}

order Taylor series expansion, we obtain:

$$\begin{aligned} b_3^0 &= b_3 + \frac{\partial b_3}{\partial p}\Delta p + \frac{\partial b_3}{\partial q}\Delta q \\ b_4^0 &= b_4 + \frac{\partial b_4}{\partial p}\Delta p + \frac{\partial b_4}{\partial q}\Delta q \end{aligned} \tag{12}$$

From (10) we can derive

$$\begin{aligned} \frac{\partial b_i}{\partial p} &= p\frac{\partial b_{i-1}}{\partial p} + q\frac{\partial b_{i-2}}{\partial p} \\ \frac{\partial b_i}{\partial q} &= p\frac{\partial b_{i-1}}{\partial q} + q\frac{\partial b_{i-2}}{\partial q}. \end{aligned} \tag{13}$$

By defining $c_1 = \frac{\partial b_2}{\partial p} = \frac{\partial b_3}{\partial q}$, $c_2 = \frac{\partial b_3}{\partial p} = \frac{\partial b_4}{\partial q}$, and $c_3 = \frac{\partial b_4}{\partial p}$, we obtain:

$$\begin{aligned} b_3^0 &= b_3 + c_2\Delta p + c_1\Delta q &= 0 \\ b_4^0 &= b_4 + c_3\Delta p + c_2\Delta q &= 0 \end{aligned} \tag{14}$$

This can be written as a system of linear equations and solved for Δp and Δq:

$$\begin{pmatrix} c_2 & c_1 \\ c_3 & c_2 \end{pmatrix} \begin{pmatrix} \Delta p \\ \Delta q \end{pmatrix} = \begin{pmatrix} -b_3 \\ -b_4 \end{pmatrix} \tag{15}$$

The algorithm is working as an iteration until $|b_3|$ and $|b_4|$ are below a defined threshold ϵ_{thr}:

1. start values: $p^{(0)} = 2cos\phi,\ q^{(0)} = -1$
   $b_0 = a_0\ \ (const.)$
2. $$\begin{aligned} b_1 &= a_1 + p^{(n)}b_0 \\ b_2 &= a_2 + p^{(n)}b_1 + q^{(n)}b_0 \\ b_3 &= a_3 + p^{(n)}b_2 + q^{(n)}b_1 \\ b_4 &= a_4 + p^{(n)}b_3 + q^{(n)}b_2 \\ c_1 &= b_1 + p^{(n)}b_0 \\ c_2 &= b_2 + p^{(n)}c_1 + q^{(n)}c_0 \\ c_3 &= b_3 + p^{(n)}c_2 + q^{(n)}c_1 \end{aligned}$$

   while $|b_3| \geq \epsilon_{thr}$ or $|b_4| \geq \epsilon_{thr}$, we get from (15):

3. $$\begin{aligned} \Delta p &= \tfrac{c_1 b_4 - c_2 b_3}{c_2^2 - c_1 c_3} \\ \Delta q &= \tfrac{c_3 b_3 - c_2 b_4}{c_2^2 - c_1 c_3} \\ p^{(n+1)} &= p^{(n)} + \Delta p \\ q^{(n+1)} &= q^{(n)} + \Delta q \\ n &= n + 1, \end{aligned}$$
   goto 2.

2.4. Discussion

To obtain a good separation between speech formant poles and oscillation poles the frequency threshold for oscillation detection should be greater than 2 kHz. Problems can arise in unvoiced speech phrases, because here the speech signal behaves like noise and this can cause wrong detections in some cases. The same is true for noisy speech signals. To circumvent this drawback an energy threshold can be introduced. Only if the signal energy per speech frame exceeds this threshold, the algorithm is calculated.

In general there is no mathematical proof, that the distance between two LSF zeros is a measure for the quality of the corresponding LPC pole. However, heuristic experiments show, that a small distance between two adjacent LSF zeros in combination with a frequency threshold is not a necessary, but most of the time a sufficient decision criterion for the detection of evolving oscillations in speech signals. If the detection criteria are relaxed and some wrong detections (sine oscillations detected, when no oscillation is present) are allowed, the algorithm gives also satisfying results by just evaluating the LSF root structure. In this case the mean frequency of two critical LSF roots (7) is taken as pole frequency for the notch filter and the Bairstow Algorithm is skipped, which means a further saving of computation power. If the 3dB-bandwidth of the notch filter is kept small and the time constants for driving the filter into and out of the signal path are chosen thoroughly, no audible distortions will be recognized.

3. CONCLUSION

An algorithm for the detection and cancellation of oscillations in speech signals has been introduced. By the combination of a 4^{th} order LPC analysis and a computationally efficient two-stage approach for the analysis of the LPC root structure oscillations can be detected early and reliably.

4. REFERENCES

[1] Joseph A. Maxwell and Patrick M. Zurek, “Reducing Acoustic Feedback in Hearing Aids,” *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 4, pp. 304–313, 1995. 1

[2] Hsiang-Feng Chi, Shawn X. Gao, and Sigfrid D. Soli, “A Novel Approach of Adaptive Feedback Cancellation for Hearing Aids,” in *Proc. IEEE*, 1999, pp. 195–198. 1

[3] Peter Kabal and Ravi P. Ramachandran, “The Computation of Line Spectral Frequencies Using Chebyshev Polynomials,” in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-34/6, 1986. 1, 2, 3

[4] “Digital Cellular Telecommunications System (Phase 2+), Enhanced Full Rate (EFR) Speech Transcoding, ETS 300 726, GSM 06.60, version 5.2.1,” ETSI, 1999. 1, 2

[5] J.H. Wilkinson, “Evaluation of Zeros of Ill-Conditioned Polynomials,” *Num. Math.*, vol. 1, pp. 150, 1959. 1, 2

[6] J.L. Morris, “Chapter 10: Zeros of Polynomial F,” in *Computational Methods in Elementary Numerical Analysis*, Wiley,1983. 1, 2, 3

[7] J.N. Mourjopoulos, E.D. Kyriakis-Bitzaros, and C.E. Goutis, “Theory and Real-Time Implementation of Time-Varying Digital Audio Filters,” *Journal of Audio Engineering Society*, vol. 38, no. 7/8, pp. 523–535, 1990. 2