Source Separation for Feedback Cancellation in Hearing Aids

Guilin Ma

Master thesis at Oersted, DTU
Supervisor: Finn T.Agerkvist & Jim Luther (GN Resound)

15th September, 2006
Abstract

This thesis addresses the problem of feedback cancellation in hearing aids using source separation techniques. Feedback oscillation (when a hearing aid whistles) is a major issue with hearing aids. It limits the maximum gain that can be achieved. The biggest barrier in solving this problem is the bias in feedback cancellation. A great number of algorithms are proposed during the last two decades. However, even in the most advanced hearing aids today, it still cannot be banished completely.

This thesis attacks the feedback problem from a new angle. It is the first attempt to use source separation techniques in feedback cancellation system. Existing algorithms on feedback suppression and source separation were reviewed thoroughly to find the fundamental limitation of feedback cancellation system and the main direction of selecting desirable source separation techniques.

Three different algorithms are investigated later. The first two are based on literature. The performance is not satisfactory. The last one, a novel algorithm proposed in this project, itself is source separation technique dealing with the very difficult case when the problem is under-determined and with correlated sources, which is unfortunately exactly the case when feedback cancellation is regarded as a source separation problem. The performance of the algorithm is satisfactory when the two sources have comparable variance. Two strategies are proposed to incorporate the proposed source separation technique in feedback cancellation system. The performance is evaluated with real speech and music on a standard feedback canceller and the most advanced feedback canceller invented by GN Resound. The first strategy is proved to be successful in enhancing the stability of the hearing aids on both feedback cancellers.
This thesis has been completed at section Acoustic Technology, Ørsted-DTU, as a partial fulfillment of the requirements for the Master of Science degree in Electrical Engineering.

The topic for the thesis is source separation for feedback cancellation in hearing aids and the project has been conducted in collaboration with the Danish hearing instrument company GN ReSound A/S. The project constitutes 35 ETCS points and has been completed in the period from February 15th 2006 to September 15th 2006 with one month break in between. The project is under the supervision of Finn T. Agerkvist, Ørsted-DTU and Jim Luther, GN ReSound A/S.

Part of the core codes are listed in Appendix.

Lyngby, September 2006

Guilin Ma
Acknowledgements

I would like to express my sincere acknowledgements to my supervisor Finn T.Agerkvist, Ørsted-DTU and Jim Luther, GN ReSound A/S for their suggestion, support and encouragement during the thesis. Without them this thesis would never have come into existence.

I am thankful to all staff at section Acoustic Technology, Ørsted-DTU, who has been supportive all along. I am also grateful to all colleagues at GN Resound. They are very friendly and helpful.

Finally, I would like to thank my friends who assisted me during this thesis. I would like to thank my parents. It is them that brought me up and give me endless love.
Table of Contents

Source Separation for Feedback Cancellation in Hearing Aids............................... i
Abstract....................................................................................................................... i
 Preface.......................................................................................................................... iii
Acknowledgements......................................................................................................... v
Table of Contents............................................................................................................. vii
List of Figures and Tables.................................................................................................. ix
Chapter 1 Introduction and Motivation........................................................................ 1
  1.1 Background............................................................................................................... 1
  1.2 Objectives.................................................................................................................. 2
  1.3 Contribution of the thesis........................................................................................ 3
  1.4 Outline of the thesis.................................................................................................. 3
Chapter 2 Feedback Suppression....................................................................................... 5
  2.1 The Feedback Mechanism....................................................................................... 5
  2.2 Feedback Suppression System.................................................................................. 7
    2.2.1 Forward path suppression.................................................................................... 7
    2.2.2 Feedback Cancellation....................................................................................... 7
  2.3 Bias in Feedback Cancellation.................................................................................. 9
  2.4 Bias Reduction in Feedback Cancellation............................................................... 12
  2.5 State of the Art in Feedback Cancellation............................................................... 15
  2.6 Performance Measures for Feedback Cancellation................................................ 16
  2.7 Summary.................................................................................................................. 17
Chapter 3 Source Separation............................................................................................. 19
  3.1 Overview of Source Separation............................................................................... 19
  3.2 Blind Source Separation.......................................................................................... 21
    3.2.1 History of blind source separation....................................................................... 21
    3.2.2 Independent Components Analysis...................................................................... 21
    3.2.3 Blind source separation with under-determined case............................................. 24
    3.2.4 Blind source separation with correlated sources...................................................... 25
  3.3 Semi-Blind Source Separation.................................................................................. 25
    3.3.1 Semi-blind source separation with under-determined case................................... 26
    3.3.2 Semi-blind source separation with correlated sources.......................................... 26
  3.4 Monaural Speech Separation..................................................................................... 27
  3.5 Summary.................................................................................................................. 27
Chapter 4 Source Separation for Feedback Cancellation................................................ 29
  4.1 Aim of using source separation in feedback cancellation system.......................... 29
  4.2 Usage of source separation in feedback cancellation.............................................. 29
    4.2.1 Indirect usage of source separation techniques in feedback cancellation............... 30
    4.2.2 Direct usage of source separation techniques in feedback cancellation.................. 30
  4.3 ICA-based LMS: increase the rate of convergence.................................................. 32
  4.4 Linear prediction: remove the bias.......................................................................... 35
  4.5 Weighted adaptive Wiener filter: A complicated scheme........................................ 39
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5.1 General ideas of the new algorithm</td>
<td>40</td>
</tr>
<tr>
<td>4.5.2 Complex Gaussian distribution and Gaussian statistical model</td>
<td>41</td>
</tr>
<tr>
<td>4.5.3 Training phase: GMM and EM algorithm</td>
<td>48</td>
</tr>
<tr>
<td>4.5.4 Separation phase: Weighted adaptive Wiener filtering</td>
<td>53</td>
</tr>
<tr>
<td>4.5.5 Performance of proposed source separation algorithm</td>
<td>56</td>
</tr>
<tr>
<td>4.5.6 Proposed algorithm in DCT domain</td>
<td>65</td>
</tr>
<tr>
<td>4.5.7 Incorporate source separation into feedback cancellation</td>
<td>66</td>
</tr>
<tr>
<td>4.5.8 Simulations results</td>
<td>67</td>
</tr>
<tr>
<td>4.6 Summary</td>
<td>77</td>
</tr>
<tr>
<td>Chapter 5 Conclusion</td>
<td>79</td>
</tr>
<tr>
<td>5.1 Work done in this thesis</td>
<td>79</td>
</tr>
<tr>
<td>5.2 Future research</td>
<td>80</td>
</tr>
<tr>
<td>Bibliography</td>
<td>83</td>
</tr>
</tbody>
</table>
List of Figures and Tables

Figure 1-1 The anatomy of ear.................................................................1
Figure 1-2 The old hearing aids ..............................................................2
Figure 2-1 Mechanism of acoustic feedback ...........................................5
Figure 2-2 Block diagram of a generic hearing aid without feedback suppression system ....6
Figure 2-3 Block diagram of a generic feedback cancellation system ............8
Figure 2-4 Compare between real feedback path and modeled feedback path using simple LMS algorithm when input signal is tonal ........................................................11
Figure 2-5 Compare between real feedback path and modeled feedback path using constrained LMS algorithm when input signal is tonal ..........................................................14
Figure 2-6 The feedback cancellation system used by GN Resound ............15
Figure 3-1 Source separation/extraction schema ....................................19
Figure 3-2 Laplacian distribution ...........................................................22
Figure 4-1 The diagram of adaptive noise cancelling problem ..................32
Figure 4-2 The compare between NLMS and Sign LMS when the input is Laplacian noise ....34
Figure 4-3 The compare between NLMS and Sign LMS when the input is male speech ....34
Figure 4-4 The compare between NLMS and Sign LMS when the input is male speech and the delay is very short .................................................................35
Figure 4-5 The diagram of feedback cancellation with linear prediction .......36
Figure 4-6 Optimal forgetting factor for clean speech and noisy speech in linear prediction ........37
Figure 4-7 Performance compare between standard feedback canceller with and without linear prediction .................................................................38
Figure 4-8 Analysis of failure reason of standard feedback canceller with linear prediction 39
Figure 4-9 The diagram of feedback cancellation with source separation .......40
Figure 4-10 Validation of assumptions for a 10th order AR process ............46
Figure 4-11 Validation of assumptions for two stationary processes ............47
Figure 4-12 An example of severe violation of assumptions for two stationary processes ..........48
Figure 4-13 General diagram of proposed source separation algorithm .........56
Figure 4-14 Results of proposed source separation algorithm when the sources are two correlated Gaussians (1) .................................................................57
Figure 4-15 Results of proposed source separation algorithm when the sources are two correlated Gaussians (2) ........................................................................58
Figure 4-16 Results of proposed source separation algorithm when the sources are two correlated Gaussians (3) ........................................................................58
Figure 4-17 Results of proposed source separation algorithm when the sources are correlated Gaussian Mixture Models .................................................60
Figure 4-18 Results of proposed source separation algorithm when the sources are two 10th AR processes ..................................................................................61
Figure 4-19 Results of proposed source separation algorithm when the sources are male speech and filtered speech ........................................................................63
Figure 4-20 Normalized test error as a function of the number of Gaussian components ..........64
Figure 4-21 Normalized test error as a function of the frame length ..............65
List of Figures and Tables

Figure 4-22 Frequency response of feedback path over time ............................................................69
Figure 4-23 Example showing how the feedback canceller behaves with different step sizes ........70
Figure 4-24 Performance evaluation of strategy I in GN Resound’s feedback canceller (flute input) ............................................................................................................................................................70
Figure 4-25 Compare between real feedback signal and estimations from feedback canceller and source separation ............................................................................................................................................................71
Figure 4-26 Performance evaluation of strategy I in GN Resound’s feedback canceller (speech input) ............................................................................................................................................................72
Figure 4-27 Performance evaluation of strategy II in GN Resound’s feedback canceller (flute input) ............................................................................................................................................................73
Figure 4-28 Performance evaluation of strategy II in GN Resound’s feedback canceller (speech input) ............................................................................................................................................................74
Figure 4-29 Performance evaluation of strategy I in standard feedback canceller (flute input)........75
Figure 4-30 Performance evaluation of strategy I in standard feedback canceller (speech input) ....76
Figure 4-31 Performance evaluation of strategy II in standard feedback canceller (flute input) ......76
Figure 4-32 Performance evaluation of strategy II in standard feedback canceller (speech input) ...77

Table 4-1 Possibility of using general source separation techniques for feedback cancellation ......31
Table 4-2 Summary of RLS algorithm for Linear Prediction ..................................................................36
Table 4-3 Test setup for standard feedback canceller and linear prediction ........................................37
Table 4-4 Random processes in the frequency domain ......................................................................43
Table 4-5 Simulation setup for source separation algorithm .................................................................57
Table 4-6 Performance as a function of the number of Gaussian components .................................64
Table 4-7 Performance as a function of frame length ...........................................................................65
Table 4-8 Simulation setup for proposed strategies ............................................................................67
Table 4-9 Optimal parameters for GN Resound’s feedback canceller .................................................69
Table 4-10 Optimal parameters for standard feedback canceller .........................................................74
Chapter 1 Introduction and Motivation

1.1 Background

The demographics of hearing loss show that an estimated 500 million people experience hearing loss worldwide today [1], which amounts to 10% of the total population.

The anatomy of a human ear which consists of three parts: outer, middle and inner ear shown in Figure 1-1 [2]. The outer ear consists of pinna and ear canal. The middle ear consists of ear drum, malleus, incus and stapes. The inner ear consists of cochlea and nerves.

Hearing loss, caused by damage in different parts of the ear, falls into two types: conductive hearing loss and perceptive hearing loss. The first type is normally attributed to problems in the middle ear. The second results from cochlear problems (in the inner ear) caused by age and/or noise, or retro-cochlea problems (after inner ear, in the nerve system from cochlea to the brain). Around 90% of hearing loss is due to the damage of hair cells in the cochlea.

Hearing loss is not only a reduced sensitivity shown as an elevated threshold in the audiogram, but a problem usually associated with abnormal loudness growth (recruitment), reduced frequency selectivity (broader critical bands and more masking), reduced temporal resolution and reduced binaural processing. Thus, hearing aid, an instrument helping hearing impaired people hear better, is far more than an amplifier. It aims at improving speech intelligibility.
Chapter 1. Introduction and Motivation

The history of hearing aids can be traced back to the 19th century. They began crudely as large trumpets and horns shown in Figure 1-2. Later, they became transistor radio style devices (analog devices) but were still bulky and uncomfortable to wear. During the mid-1980s the first digital hearing aids were launched, but these early models were slightly unpractical. It was not until ten years later that digital hearing aids really became successful, with small digital devices placed either inside or discreetly behind the ear. Some modern hearing aids are even implantable and therefore invisible. [3]

Modern hearing aids are proved to be helpful to those hearing-impaired people. However, no more than 40% of those people over 65 actually use hearing aids. Besides the cost factor, comfort is another important issue. This discomfort includes the following cases: the wearers could feel the sounds unnatural. They may encounter problems with understanding in a party. They can also hear their own voice very loud due to the occlusion effect. They often experience loud whistles when picking up the phone or putting hands close to the hearing aids. Among all the uncomfortable problems, customers especially complain about the whistling problems, i.e. the feedback problem according to the information GN Resound has gathered.

The sound output from the receiver can be picked up by the microphone in the hearing aids. If the amount of amplification through the hearing aids is greater than the amount of attenuation from the ear canal back to the microphone, oscillation occurs. In addition, the requirement for phase is that the total delay around the entire loop must be an integer number of periods of the feedback signal. [4] However, in practice, the phase requirement is negligible.

1.2 Objectives

Feedback reduction in hearing aids has been studied intensively over the last 20 years. Many techniques have been developed to address the problem and proved to be helpful. However, none of them can banish feedback oscillation entirely.

The aim of this project is trying to investigate if source separation techniques can be helpful in the feedback cancellation system of modern hearing aids. Being aware of the immaturity in source...
1.3 Contribution of the thesis

separation today, we only aim at using source separation to enhance the existing feedback cancellation system instead of replacing.

This is a brand new idea. Thus this project is basically a tentative research. It involves several parts: make a survey of existing feedback suppression techniques and find out the limitations of performance; review the source separation problems and try to find a suitable method or derive new methods of source separation for feedback cancellation; figure out a way for source separation to enhance feedback cancellation system; do simulation to verify the design.

1.3 Contribution of the thesis

This thesis is the first attempt to use source separation techniques for feedback cancellation in hearing aids. Contributions are made on the following aspects: A thorough review of feedback suppression techniques is made and the state of art in feedback cancellation system is elaborated. A comprehensive review of various source separation techniques is presented by collecting and categorizing techniques scattered in papers and books. Such a review is rarely seen in literatures. A novel algorithm is proposed for the difficult monaural source separation when the two sources are correlated. The algorithm is proved be very effective and the way of reducing computation load is also described. Two new strategies to incorporate the new algorithm in feedback cancellation systems are proposed. With the first strategy, source separation assisted feedback canceller is proved to be superior in the sense of stability of hearing aids.

1.4 Outline of the thesis

The remaining chapters of this thesis are organized as follows: In Chapter 2, existing feedback suppression techniques are reviewed to provide a basic understanding of the feedback problem and to find out the fundamental issue that limits the performance. In Chapter 3, a broad review of various source separation techniques is presented to help with selection of proper candidates for enhancing feedback cancellation system in the later stage of the project. In Chapter 4, the possibility of utilizing source separation in feedback cancellers is investigated. Two proposals are investigated and one novel approach is proposed and tested with real speech and music. In Chapter 5, the work having been done in this project is summarized and the future research is pointed out.
Chapter 2 Feedback Suppression

In this Chapter, feedback problem in hearing aids is investigated. An overview of existing feedback suppression techniques is given to provide a basic understanding of the problem and to find out the fundamental issue that limits the performance.

This chapter is organized as follows: Section 2.1 describes the mechanism of feedback problem from two respects: physical reason and steady-state analysis of the hearing-aid system. Section 2.2 introduces the general feedback suppression techniques which are divided into two types: forward path suppression and feedback cancellation. The bias problem is investigated by theoretical deductions and simulations in section 2.3. After that, section 2.4 gives a review of various bias reduction techniques. Section 2.5 elaborates the most advance technique in feedback cancellation. Section 2.6 deals with the methods to evaluate the performance of feedback cancellers. In the end, section 2.7 summarizes this chapter.

2.1 The Feedback Mechanism

Feedback oscillation (when a hearing aid whistles) is a major problem with hearing aids. It limits the maximum gain that can be achieved in most hearing aids. Without effective feedback suppression the maximum stable gain for a vented behind-the-ear (BTE) or in-the-ear (ITE) hearing aid is limited to 40 dB for small vents and even less for larger vents or the new open ear mould type hearing aids. [5][6]

The term feedback literally means that some of the output of the hearing aids manages to get back to the input of the hearing aid (i.e., it is fed back to the input). [4] In hearing aids, two kinds of feedback exist: mechanical feedback and acoustic feedback. The mechanical feedback is usually caused by the vibrations of the receiver diaphragm being transmitted back to the microphone diaphragm through contact with the hearing-aid shell. The acoustic feedback shown in Figure 2-1 includes the effects of the hearing-aid amplifier, receiver, and microphone as well as the acoustics of the vent or leak.

![Figure 2-1 Mechanism of acoustic feedback](image)
To illustrate that feedback can lead to instability of the system and thus results in whistle, the equivalent diagram to Figure 2-1 is shown below:

![Figure 2-2 Block diagram of a generic hearing aid without feedback suppression system](image)

The input to the hearing-aid processing is $s(n)$, which is the sum of desired input signal $x(n)$ and the feedback signal $f(n)$. The processed hearing-aid signal is $u(n)$. The amplifier impulse response is given by $a(n)$, the receiver impulse response by $r(n)$ and the microphone impulse response by $m(n)$. The signal in the ear canal is $y(n)$. The feedback path impulse response $b(n)$ includes both the acoustic (mainly contributed by a vent leading from the ear canal back to the microphone) and mechanical feedback, although acoustic feedback is assumed to dominate. The acoustic feedback path through the vent tends to have a high-pass behavior, and the un-amplified acoustic feed-forward path $c(n)$ through the vent from pinna to the ear canal tends to have a low-pass filter behavior. [7]

[7] gives a complete steady-state analysis of generic feedback cancellation system. The following results are based on the analysis but removing the feedback cancellation part to show how feedback triggers the whistle. The results are presented in the frequency domain, denoted by upper-case variables (a convention adopted all over this thesis), and the frequency variable is $f$. The output signal is given by:

$$Y(f) = \frac{X(f)[C(f) + H(f)M(f)A(f)R(f)]}{[1-B(f)\cdot C(f)]-H(f)M(f)A(f)R(f)B(f)}$$  (2.1)

Because $C(f)$ is a low-pass response and $B(f)$ is a high-pass response with reduced gain, one can assume safely that the product that the product $|B(f)C(f)| << 1$. This assumption leads to a useful approximation solution:

$$Y(f) \approx \frac{X(f)[C(f) + H(f)M(f)A(f)R(f)]}{1-H(f)M(f)A(f)R(f)B(f)}$$  (2.2)

As the denominator of Eq. (2.2) deviates from 1, the system transfer function deviates from the desired response. The danger occurs for a denominator close to zero at some frequency; if the denominator actually reaches zero, the transfer function for $Y(f)$ involves a division by zero which indicates infinite output amplitude. A stable system is defined as one in which a finite input yields a finite output, therefore the hearing-aid system becomes unstable. In practice an infinite output level can not be achieved given the finite power stored in the hearing-aid battery, but the system tries its best to reinforce the output at the frequency where the gain is infinite, resulting in a continuous
whistle output by the hearing aid. [8] A mathematical statement is that stability is ensured if the hearing-aid gain $H(f)$ is selected so that:
\[
[H(f)[M(f)A(f)R(f)B(f)] < 1
\] (2.3)
at all angular frequencies $f \in [0,1/2]$ with positive feedback, i.e. [8][9]:
\[
\angle H(f)[M(f)A(f)R(f)B(f)] = n \cdot 2\pi, n \in Z. 
\] (2.4)

### 2.2 Feedback Suppression System

To address the feedback problem, many schemes have been proposed in the last two decades. In this section, these methods will be reviewed. The existing techniques can be classified into forward path suppression and feedback cancellation.

#### 2.2.1 Forward path suppression

Forward path suppression techniques modify the forward path, $H(f)$, of the hearing aids in such a way that it is stable in conjunction with feedback path $B(f)$.

The simplest solution to reduce the occurrence of feedback problems is to reduce the gain in order to satisfy Eq (2.3). But the gain does not need to be reduced everywhere; it only needs to be reduced in the critical frequency region where feedback problems are expected to occur.[8] This leads to a notch or adaptive notch filter, or gain reduction scheme with a whistle detector. This technique is useful to increase the maximum gain at other frequencies other than the critical frequencies. However, the closed-loop gains are still limited and frequency response of the hearing aid can be severely compromised.

The other ways of forward path suppression include delay modulation, phase modulation and frequency shifting, modifying the phase of open-loop response of $H(f)M(f)A(f)R(f)B(f)$ in (2.2) so that the build-up of feedback resonances is disrupted.[9]

In summary, though the forward path suppression helps to some extent, the performance is not satisfied. It either compromises the sound quality or limits the gain in critical region which contradicts with the amplification function of hearing aids.

#### 2.2.2 Feedback Cancellation

A more promising way of feedback suppression is feedback cancellation as illustrated in Figure 2-3. The idea is to include an adaptive module to model the feedback path and subtract the modeled response from the signal input to the hearing aids. A perfect match between modeling and real feedback path will completely cancel the feedback signal $f(n)$ and the system will be completely stable for any amount of amplification. In practice, feedback cancellation systems typically permit 10 to 15 dB additional gain in hearing aids before the system becomes unstable, although the system benefit is often less for tonal inputs or for large change in the feedback path.
Chapter 2. Feedback Suppression

The steady-state analysis is given in [7]:

$$ Y = \frac{X(f)\left[C(f) + H(f)(W(f)C(f) + M(f)A(f)R(f))\right]}{(1 - B(f)C(f)) - H(f)[M(f)A(f)R(f)B(f) - W(f)(1 - B(f)C(f))]} $$

(2.5)

By assuming $|B(f)C(f)| << 1$, (2.5) is reduced to:

$$ Y \approx \frac{X(f)\left[C(f) + H(f)(W(f)C(f) + M(f)A(f)R(f))\right]}{1 - H(f)[M(f)A(f)R(f)B(f) - W(f)]} $$

The stability is guaranteed by the Nyquist criterion if

$$ |H(f)[M(f)A(f)R(f)B(f) - W(f)]| < 1 $$

(2.6)

If $\hat{W}(f)$ is a good model of the actual feedback path $M(f)A(f)R(f)B(f)$, Eq (2.6) will be satisfied. On the contrary, if there is much difference between them, the maximum stable gains of $H(f)$ will be reduced a lot. Moreover, the output sound quality will be degraded.

Many schemes have been tried out for the adaptive module in the feedback cancellation systems. Currently, the available feedback cancellers can be divided into two classes: algorithms with continuous adaptation and algorithms with a non-continuous adaptation. [9] The latter adapt only when instability is detected or the input signal is proper. This kind of techniques may be objectionable due to the reactive rather than proactive adaptation. [9]

A continuous adaptation feedback canceller (CAF) continuously adapts the coefficients of the filter, based on standard adaptive filtering procedures, see [10]. However, due to the requirement of low-complexity for algorithms in hearing aids, only simple adaptation is used. The simplest form of adaptation is the LMS (least mean square) algorithm. The adaptive FIR (finite impulse response) filter coefficients are adapted by minimizing the cost function:

$$ J(n) = \|e(n)\|^2 $$

(2.7)

The update formula is given below (vector or matrix is denoted by bold all over this thesis):

$$ \hat{w}(n+1) = \hat{w}(n) + \mu(n) \cdot u(n) \cdot e^*(n) $$

(2.8)

where, $\mu(n)$ is the step-size, which could be time-varying, $\hat{w}(n)$ is the coefficients of adaptive filter in the form of $\hat{w}(n) = [\hat{w}_1(n) \quad \hat{w}_2(n) \quad \cdots \quad \hat{w}_{p_1}(n)]^T$, $p$ is the order of FIR filter, $u(n)$ is the input
vector shown in Figure 2-3 in the form of $u(n) = [u(n) \quad u(n-1) \cdots u(n-p+1)]^T$ and $e(n)$ is the error signal shown in Figure 2-3.

With a wide sense stationary input and a proper step size, the adaptive filter will converge towards the Wiener solution. Ideally, the adaptive filter coefficients $\hat{w}$ converge to the real feedback path response. So that, the error signal $e(n)$ equals the desired signal $x(n)$ which is different from zero. Hence, to limit the excess MSE (mean square error), the step size should be kept small, at the expense of slow convergence. If the desired sound signal $x(n)$ exhibits large fluctuations in the short-time power (e.g., if $x(n)$ is a speech signal), the excess MSE of the LMS algorithm is further reduced by using a normalized time-varying step size based on the sum-method, which is also referred to as the modified LMS algorithm [9]:

$$
\mu(n) = \frac{\bar{\mu}}{p(p_u(n) + p_e(n)) + c}
$$

where $\bar{\mu} \in (0,2]$ is a step size constant, $c$ is a positive constant to prevent singularities, $p_u(n)$ and $p_e(n)$ are running power estimate of $u(n)$ and $e(n)$ respectively:

$$
p_u(n) = \gamma p_u(n-1) + (1-\gamma)u^2(n)
$$

$$
p_e(n) = \gamma p_e(n-1) + (1-\gamma)e^2(n)
$$

with the forgetting factor $\gamma$. Eq. (2.9) shows that when the desired signal $x(n)$ is strong the step size will be reduced and thus the negative effect of a strong desired signal on the excess MSE is mitigated.

### 2.3 Bias in Feedback Cancellation

As mentioned above, the adaptive filter will converge towards the Wiener solution and ideally the adaptive filter coefficients $\hat{w}$ converge to the real feedback path response. However, the Wiener solution is not necessarily close to the real feedback path response even when the order of the filter is sufficiently high and quantization error of the filter coefficients is negligible. Actually, the adaptive filter works best when the input signal $x(n)$ is white Gaussian noise. When the input and output of the hearing aids are correlated, the Wiener solution becomes a biased estimation of a feedback path. To show this bias effect, a theoretical formulation and a simulation result are given below.

The cross-correlation between $s(n)$ and $u(n)$ is given by:

$$
\gamma_{su}(i) = E[s(n)u(n-i)] = E[(x(n) + f(n))\cdot u(n-i))] = \gamma_{su}(i) + \gamma_{fu}(i)
$$

(2.12)

The Wiener solution that minimizes the error signal $e(n)$ is given by:

$$
\hat{w} = \Gamma_{uu}^{-1}\gamma_{su} = \Gamma_{uu}^{-1}(\gamma_{su} + \gamma_{fu})
$$

(2.13)

where $\Gamma_{uu} = E[u(n)u^T(n)]$, $\gamma_s = [\gamma_s(0), \gamma_s(1), \ldots, \gamma_s(p-1)]^T$. However, assuming that the microphone response is ideal and the un-amplified feed-forward path $c(n)$ is negligible, the real feedback path is given by:

$$
\hat{w} = \Gamma_{uu}^{-1}\gamma_{fu}
$$

(2.14)

Thus,

$$
\hat{w} = w + \Gamma_{uu}^{-1}\gamma_{su}
$$

(2.15)
The term $\mathbf{I}^{-1}_{uu} \mathbf{y}_{uu}$ is the bias. If the desired signal $x(n)$ is zero-mean and statistically independent of the input vector $u(n)$, i.e., $E[u(n)x(n)] = 0$, the estimate of $\hat{w}$ is unbiased. Otherwise, the cross-correlation can be rewritten as:

$$\gamma_{xx}(i) = \gamma_{ux} = h(-i) \ast \gamma_{uu}(-i) = h(-i) \ast \gamma_{ux}(i) = h(0) \gamma_{xx}(i) + h(1) \gamma_{xx}(i+1) + \cdots$$

(2.16)

Eq. (2.16) shows that the hearing-aid process $h(n)$ affects the cross-correlation and the more auto-correlated the input signal is the more cross-correlation exists between $u(n)$ and $x(n)$. With a white noise, the cross-correlation is zero due to the delay in the hearing-aid process and thus the estimation is unbiased. However, with a tonal or narrow band input the cross-correlation between $u(n)$ and $x(n)$ may be the dominant term in (2.13) and hence the Wiener filter becomes a predictor of $x(n)$ instead of an estimate of the feedback path. In this way, the feedback cancellation system will try to cancel the desired signal instead of feedback signal and degrade the sound quality very much.

The introduced bias severely limits the performance of the feedback cancellation system especially when tonal signal or music signals are input to the hearing aids. This is shown by a Matlab simulation. In the simulation responses of microphone, receiver, amplifier and the un-amplified acoustic feed-forward path are simply ignored. The hearing-aid process $h(n)$ is a delay of 8 samples and the gain is 0 dB. The input signal is a 1 kHz sinusoid with amplitude 0.4. The sampling frequency is 16 kHz. LMS algorithm is used with the step size of 0.0005. Both the real feedback path and the adaptive filter are 120-order. The code is shown in the appendix and the results are shown below:
2.3 Bias in Feedback Cancellation

From Figure 2-4 it can be seen that there is big mismatch between the real feedback path and modeling. The big mismatch is understandable if we follow the signal flow in the simulation carefully.

When a sinusoid $x(n)$ is input to the hearing aid as, it will also be present in the error signal $e(n)$. The sinusoid will be amplified by the hearing aid processing, so an amplified and delayed version of the sinusoid will also appear in the hearing-aid output $y(n)$ and in the input to the adaptive filter $u(n)$. Because $e(n)$ and $u(n)$ are both sinusoids at the same frequency, they will be highly correlated, and the correlation between the two sinusoids will be much greater than any signal correlations related to the feedback path transfer function. Because the adaptive filter coefficient update is driven to minimize the error signal, the filter will adapt by shifting the amplitude and phase of $u(n)$ so that the filtered signal $v(n)$ cancels the sinusoid at the microphone output $s(n)$. The result is that for a sinusoidal input signal, adaptive feedback cancellation stops modeling the feedback path and instead adapts to cancel the input signal. [8] When the microphone signal is heavily distorted, the cancelling effect will be smaller since the cross-correlation is diminished and as a result, the microphone signal increases again. This explains why in the figure the output signal from hearing aids is fluctuating. Figure 2-4 also shows a noticeable characteristic of the coefficients of the adaptive filter. The magnitude of the adaptive FIR filter coefficients are normally very large and drifted far from the real feedback path when the bias severely influences the adaptation. The excessively large filter coefficients, when combined with the hearing-aid gain function, can result in system instability or coloration of the output signal due to large undesired changes in the system frequency response. This observation gives rise of a technique named constrained adaptation to reduce bias which will be discussed later.
2.4 Bias Reduction in Feedback Cancellation

Bias has been the main issue and performance limitation in the feedback cancellation system. So far, much discussion has focused on this topic, but no satisfied solution has been found yet.

Adding delay in the forward path One way to address the bias problem is including a delay in the forward path since it helps reducing the cross-correlation between $x(n)$ and $u(n)$. In order to efficiently de-correlate $x(n)$ and $u(n)$, a large delay is required, which will degrade intelligibility and sound quality. In addition, adding delay will not help if $x(n)$ is tonal as is the case for voiced speech, music, alarm whistling signals. [9]

Nonlinearities in the forward path A variant of de-correlation in the forward path is to add nonlinearities (e.g., frequency shifting, phase or delay modulation) instead of delay. However, these methods again degrade sound quality and the increased gain is very limited.

Adding probe signal in output Adding probe signal, which generally is a white noise, to the output can increase the cross-correlation between $x(n)$ and $f(n)$ and keep the cross-correlation between $x(n)$ and $u(n)$ unchanged. This will lead to a better estimate of feedback path as seen from (2.13). However, in order to maintain the output sound quality, the level of the probe signal should be inaudible. Thus, significant residual bias still remains. Another way of utilizing probe signal is to use non-continuous probe signal during continuous adaptation which is described in [12].

Controlling the adaptation of the feedback canceller The bias of the adaptive feedback canceller can also be reduced by controlling the adaptation. One way is to changing the speed of adaptation. When a sudden change in the feedback path is detected or when the input signal is white noise, the adaptation speed is increased. Otherwise, the adaptation speed is kept small to reduce distortion of the desired signal. The second way is to restrict feedback cancellation to the frequency band that encompasses the unstable frequencies by using a filtered-X technique [9], in which $u(n)$ and $e(n)$ are filtered by the same filter, limiting the feedback cancellation signal to the frequency band where possible oscillation may occur, before going into the adaptation algorithm. A filtered-X LMS based feedback canceller with adaptive de-correlation filters has also been proposed in [13]: $u(n)$ and $e(n)$ are pre-whitened through adaptive linear prediction filters before updating the feedback canceller. Linear prediction removes the correlation between consecutive samples of a signal and hence reduces the autocorrelation function. It helps to reduce the bias of the feedback canceller. Another way of controlling the adaptation is called constrained adaptation which has been put into commercial hearing aids of GN Resound. It is elaborated below.

Constrained feedback canceller The constrained adaptation aims at modeling the feedback path accurately and avoids cancelling a sinusoid or tonal input. To achieve this end, the adaptation proceed normally as long as the adaptive filter remains close to an assumed “correct” reference filter, and prevents the adaptive filter coefficients from drifting too far from this presumed correct solution. This technique needs an additional initialization of feedback canceller during the fitting of hearing aids to provide a reasonably good reference since in practice there is no “correct solution” available. In the initialization a noise-like periodic excitation sequence is used to measure the feedback path response of the hearing aid inserted in the user’s ear. As mentioned above, the adaptive filter works best when the input signal is a white noise. The normal hearing-aid processing is disabled during the initialization, so the system measures the feedback path response in situ, and produces an accurate estimate of the overall transfer function $M(f)A(f)R(f)B(f)$. A response is
collected for each period of the excitation, and the responses are averaged together to suppress the effects of external noise sources. Because the excitation is periodic, it creates a buzzing sound in the listener’s ear while the initialization is performed. [8] The initial model of feedback path is denoted as \( \hat{w}(0) \) in time domain and \( \hat{F}(f,0) \) in the frequency domain, which will later serve as the reference filter. The stability criterion (2.6) is approximated by \( |H(\hat{F}(f,0) - \hat{F}(f,n))| < 1 \), where \( \hat{F}(f,n) \) is feedback path model for processing block \( n \), assuming that the feedback path has not changed substantially. Stability for a sinusoidal input will be maintained as long as the feedback path model stays close to the initial measurement. [8]

Two types of constraints were proposed by Kates [14]. The first method is called clamp, which allows the coefficients to freely adapt within a distance from the reference filter. The boundary is hard and the coefficients cannot go beyond it. The coefficients are not updated unless the following requirement is met:

\[
\frac{\sum_{k=0}^{p-1} |\hat{w}_k(n) - \hat{w}_k(0)|}{\sum_{k=0}^{p-1} |\hat{w}_k(0)|} < \gamma \tag{2.17}
\]

where \( \gamma \approx 2 \) gives the desired 10 dB headroom above the reference condition to tolerate the feedback response change when a telephone is placed near the aided ear.

The second constraint modifies the error given by (2.7) that is used to guide the coefficient adaptation. The modification is a cost constraint that penalizes deviations of the adaptive filter coefficients if they move away from the initial values. The modified error contains a penalty for large filter coefficients:

\[
J(n) = |e(n)|^2 + \beta \sum_{k=0}^{p-1} [\hat{w}_k(n) - \hat{w}_k(0)]^2 \tag{2.18}
\]

The LMS adaptive filter coefficient update equation then becomes:

\[
\hat{w}(n+1) = \hat{w}(n) + \mu(n) \cdot e(n) \cdot e^*(n) - \mu(n) \beta [\hat{w}(n) - \hat{w}(0)] \tag{2.19}
\]

When \( \mu(n) \) is a constant, it is based on a simple LMS method. When \( \mu(n) \) is updated according to (2.9), it is based on a modified LMS algorithm.

The constraints prevent the system from canceling a sinusoidal input, but may also prevent the system from adapting fully to a large change in the feedback path. [8]

A computer simulation was made to show how the cost function constraint works. The setup of the simulation is the same as in Figure 2-4. But the adaptation is constrained LMS with the step size \( \mu(n) = 0.0005, \beta = 0.0005 \). The reference filter \( \hat{w}(0) \) is set as the real feedback path for simplicity. The code is shown in the appendix and the results are shown below:
Compared with Figure 2-4, the effect of cancelling the input signal is obviously weakened. The modeled feedback response is not drifted far away from the real feedback path due to the constraint. However, because of the remaining cancelling effect, the output of hearing aids is a sinusoid with the same frequency as input but smaller amplitude though the gain of the hearing-aid process is 0 dB.
2.5 State of the Art in Feedback Cancellation

GN Resound has taken the lead in feedback canceller in the hearing-aid industry. This section will describe the principle of its feedback canceller whose effectiveness has been proved by the market.

The following diagram shows the structure of the feedback canceller:

![Feedback Cancellation System Diagram](image)

Figure 2-6 The feedback cancellation system used by GN Resound

The focus in GN Resound’s feedback canceller is to reduce sensitivity to narrowband input signals whose autocorrelation is normally very large. It consists of a fixed IIR filter, a filtered-X modified LMS adaptive filter with constraints.

The impulse response of a typical feedback path is very long and contains a bulk delay as shown in Figure 2-4 and Figure 2-5. Instead of using a long adaptive FIR filter, an IIR filter is added to model the invariant part of the feedback path including microphone, receiver, amplifier and the basic acoustical feedback path. In this way, the adaptive FIR filter can be much shorter and efficient. The constrained adaptation based on a modified LMS algorithm happens mainly in the interesting frequency range due to the two identical high-pass filters (a filtered-X technique). The delay is determined during the initialization of the feedback cancellation system to match the delay in the feedback path. The DC filter is included to remove the mismatch in low frequencies and also to comply with the fact that the receiver cannot output DC.

A typical setup of the system is like the following: The IIR has five poles and the adaptive FIR has twelve taps. The DC filter is a two-tap FIR filter with coefficients [0.667, -0.333]. The update formula is:

\[
\hat{w}_k(n+1) = \hat{w}_k(n) - \beta [\hat{w}_k(n) - \hat{w}_k(0)] + \frac{\mu}{p_r(n) + c} \sum_{m=0}^{L} e_m^f(n)v_{m-k}^f(n)
\]

where, \(L\) is the number of samples per block, \(v_f(n)\) and \(e_f(n)\) are high-pass filtered excitation and error signals respectively, and
\[ p_r(n) = (1 - \lambda) v_r(n)^2 + \lambda p_r(n-1), \lambda = 0.995, c = 10^{-5} \quad (2.21) \]

Note that in (2.20), the block-LMS is used and \( \beta \) has absorbed \( \mu \) (compare with (2.19)).

An initialization is performed during the fitting of hearing aids to determine the amount of proper delay, coefficients of IIR filter and reference adaptive FIR filter. The details of the initialization are described in [11].

2.6 Performance Measures for Feedback Cancellation

The performance of the feedback cancellation system can be assessed by several ways, among which, the misalignment between the estimated and real feedback path, the maximum or additional stable gain are most popular and will be used later as a measure for the new algorithm. More advanced ways can be found in [11] and [15].

**Misalignment**

The misalignment between the estimated feedback path \( \hat{W}(f) \) and the real feedback path \( M(f)A(f)R(f)B(f) \), denoted in the frequency domain as \( \tilde{W}(f) \) is defined as:

\[
\zeta(\hat{W}(f), W(f)) = \sqrt{\frac{\int_0^{\pi/2} |W(f) - \hat{W}(f)|^2 df}{\int_0^{\pi/2} |W(f)|^2 df}}
\]

(2.22)

It reflects the accuracy of the feedback path estimation. [9] There is also a counterpart in time domain defined as below:

\[
\zeta(\hat{w}, w) = \frac{\|\hat{h} - h\|}{\|h\|}
\]

(2.23)

where \( h \) and \( \hat{h} \) represent impulse response of real feedback path and modeled impulse response respectively.

**Maximum stable gain (MSG) and additional stable gain (ASG)**

MSG and ASG show the benefit the feedback cancellation system provides. MSG is especially widely used. MSG is defined as the maximum gain without instability assuming a flat response of the hearing-aid process and the worst case for the phase [9]:

\[
MSG = 20 \log_{10} \left( \min_f \frac{1}{|W(f) - \hat{W}(f)|} \right)
\]

(2.24)

MSG shows the frequency where the largest mismatch is located. ASG is defined as the additional gain that is possible by using the feedback canceller:

\[
ASG = MSG - 20 \log_{10} \left( \min_f \frac{1}{|W(f)|} \right)
\]
2.7 Summary

This chapter describes the mechanism of feedback problem and makes a review of feedback suppression techniques. The feedback cancellation is especially paid attention to since it has become a standard method for feedback suppression and widely used in existing hearing aids. A computer simulation is made to find out the main issue of the feedback canceller, i.e., the bias introduced due to the cross-correlation between input and output of the hearing aids. Many techniques of bias reduction are discussed, among which, constrained adaptation is elaborated since it has been put into use in commercial hearing aids. A computer simulation also proves the effectiveness of the adaptation algorithm.
Chapter 3 Source Separation

This chapter gives an overview of source separation techniques which have become an active research area recently. The aim of the overview is to get a broad picture of the topic and show clearly the use of different methods in different situations. It helps with selection of proper methods to enhance feedback cancellation system in the later stage of the project.

This chapter is organized as follows: Section 3.1 generally describes the source separation problem, its categories and the essence of solutions. Section 3.2 and 3.3 deal with blind and semi-blind source separation techniques respectively. They together give an overall picture of the techniques in source separation. Under-determined case and correlated case are paid attention to in these two sections due to their relevance to the feedback cancellation system. Section 3.4 summarizes the techniques for monaural speech separation. In the end a summary is given in section 3.5.

3.1 Overview of Source Separation

Source separation arises in a variety of signal processing applications, ranging from speech processing to medical image analysis. When presented with a set of observations from sensors such as microphones, the process of extracting the underlying sources is called source separation. [16] The schema of source separation is shown below [17]:

![Source separation/extraction schema](image)

As a very broad and fundamental problem, source separation problem can be categorized in several ways: Depending on the amount of available information about the mixing process and sources, it can be divided into blind separation and semi-blind separation; According to the relation of $n$ (the number of sources) and $m$ (the number of sensors), it falls into under-determined problems ($m < n$),
even-determined problems \((m = n)\) or over-determined problems \((m > n)\); Based on the relation between sources, it is either a problem with independent sources or a problem with correlated sources.

**Mathematic formulation of source separation problem**

The simplest model of source separation problem is the linear instantaneous noiseless mixing which can be presented as follows: assumes the existence of \(n\) unknown sources \(s(t) = (s_1(t), s_2(t), \ldots, s_n(t))^T\), and the observation of mixtures \(x(t) = (x_1(t), x_2(t), \ldots, x_m(t))^T\). Assume the unknown mixing processing is linear and instantaneous, characterized by the \(m \times n\) mixing matrix \(A\), yielding the equation:

\[
x(t) = As(t)
\]

which can also be written as:

\[
x(t) = \sum_{i=1}^{n} a_i s_i(t)
\]

The aim of source separation is to restore \(s(t)\) from observed \(x(t)\) without knowing \(A\).

**Assumptions about environment and sources**

The separation of a superposition of multiple signals is accomplished by taking into account the structure of the mixing process and by making assumptions about the sources. It is impossible to estimate the underlying sources without any a priori knowledge or assumptions. The complexity of the algorithm is mainly determined by the inherent characteristics of the mixing process and the nature of sources.

For acoustic signal separation, the mixing process is usually dependent on the environment. The complexity could be totally different when the environment changes from an anechoic room to a reverberant environment. The mixing process can even involve nonlinearity in very complicated surroundings. The relation between \(m\) and \(n\) is another very important factor. The under-determined problems are usually far more complicated than the over-determined and even-determined problems in any case. Moreover, if the mixing process includes noise, the performance of separation algorithms could be degraded a lot. As a matter of fact, although there exist a large number of source separation algorithms, every individual algorithm makes its own assumption about the mixing process and only suits the specific needs. Many algorithms depart from the difficult real-world scenario and make less realistic assumptions about the mixing process to make the problem more tractable.

Most algorithms have to make assumptions about the nature of sources since they form the basis for the algorithms. [16] The assumptions normally include the statistical properties including the relation among sources and inherent characteristics of individual source. The sources can be independent, uncorrelated or correlated between each other. The independent case is especially widely used since in real world sounds coming from different sources are mutually independent in most cases. However, in some special situations, such as for radar images, sources are correlated. The algorithms for correlated case are completely different from those for independent cases. The statistical nature of sources could be described by the stationarity, sparsity, probability distribution, Gaussianity and temporal correlation. The stationarity is usually assumed in most algorithms. The sparsity of sources in a given basis is an increasingly popular and powerful assumption. A signal is said to be sparse when it is zero or nearly zero more than might be expected from its variance. Such a signal has a probability density function or distribution values with a sharp peak at zero and far
3.2 Blind Source Separation

tails. A sparse representation of an acoustic signal can often be achieved by a transformation into a Fourier, Gabor or Wavelet basis. [16] Gaussianity is a key assumption to a big class of algorithms since “non-Gaussian is independent” in some sense. If sources are colored, temporal structures of the source have to be considered. Temporally correlated source signals can be modeled, for example, by autoregressive processes (AR).

Source separation algorithms A large variety of algorithms have been proposed for source separation. They are from different starting points, such as information theory, neural network and so on. Some of the techniques are very general, for example, independent component analysis, whereas some are very ad hoc, only aiming at solving specific problems. So far, no literature has been able to make a broad review of the large number of techniques. The following sections only try to make a survey of the main techniques.

Actually, the problem of source separation is by its very nature an inductive inference problem. There is not enough information to deduce the solution, so one must use any available information to infer the most probable solution. This information comes in two forms: the signal model and the probability assignments. By adopting a signal model appropriate for the problem, an algorithm that is specially-tailored to suit the needs can be developed. The quality of the results depends on the information put into the algorithm. An algorithm will do better when more specific knowledge is incorporated. [19]

3.2 Blind Source Separation

Blind source separation (BSS) has been studied for two decades. It is an important subject for source separation and is still a very hot research area. The strength of the BSS model is that no a priori information about, e.g., the characteristics of the source signals, the mixing matrix or the arrangement of the sensors is needed. Therefore BSS can be applied to a variety of situations such as, e.g., the separation of simultaneous speakers, analysis of biomedical signals obtained by EEG or in wireless telecommunications to separate several received signals.

3.2.1 History of blind source separation

Many people prefer blind source separation since it is a general technique requiring little information about sources and mixing process, and can be used anywhere. The earliest approach is traced back to 1986 in separating an instantaneous linear even-determined mixture of non-Gaussian independent sources. Later, many researchers devoted themselves to this subject. Independent Component Analysis (ICA) is definitely a strong tool for BSS. Some extensions of ICA were later developed including fast ICA and local ICA methods. The early years of BSS research concentrated on solutions for even-determined and over-determined mixing processes. It was not until 1994 that a solution for the under-determined case was proposed.

3.2.2 Independent Components Analysis

ICA is a recently developed method in which the goal is to find a linear representation of non-Gaussian data so that the components are statistically independent or as independent as possible.
Such a representation seems to capture the essential structure of the data in many applications, including feature extraction and signal separation. [22]

The starting point of ICA is the very simple assumption that the components $s_i, i = 1, 2, \cdots, n$ are statistically independent. The fundamental restriction in ICA is that the independent components must be non-Gaussian. (Actually, if just one of the independent components is Gaussian, the ICA model still works.)

In the ICA method, some ambiguities exist after estimation: (i) The variance of the independent components cannot be determined. Since both $s(t)$ and $A$ are unknown, any scalar multiplier in one of the sources $s_i$ could always be cancelled by dividing the corresponding column $a_i$ of $A$ by the same scalar. A natural way of fixing this ambiguity is to assume that each component has unit variance. However, this still leaves the ambiguity in the sign: an independent component multiplied by -1 will not affect the model. (ii) The order of the independent components cannot be determined. Again since both $s(t)$ and $A$ are unknown, the order of the terms can be freely changed. Thus the estimation of the mixing matrix is up to permutation and scaling of the rows.

**Maximizing non-Gaussianity**

The key to estimating the ICA model is non-Gaussianity which is based on the central limit theorem. One approach of estimating the independent components is maximizing the non-Gaussianity. To use non-Gaussianity in ICA estimation, quantitative measures of non-Gaussianity of a random variable, say $y$, have to be defined.

The classical measure of non-Gaussianity is kurtosis of the fourth-order cumulant which is defined as:

$$kurt(y) = E(y^4) - 3(E(y^2))^2$$  \hspace{1cm} (3.3)

Kurtosis is zero for a Gaussian random variable, negative for subgaussian and positive for supergaussian. In statistical literature, the corresponding expressions platykurtic and leptokurtic are also used. Supergaussian random variables have typically a “spiky” probability density function (pdf) with heavy tails. A typical example is the Laplace distribution whose pdf (normalized to unit variance) is given by

$$p(y) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|y|}$$  \hspace{1cm} (3.4)

The pdf is illustrated below:

![Figure 3-2 Laplacian distribution](image)
Subgaussian random variables, on the other hand, have typically a “flat” pdf. A typical example is the uniform distribution.

Another important measure for non-Gaussianity is given by negentropy which is based on the information theoretic quantity of entropy. Entropy is defined for a discrete random variable \( Y \) as:

\[
H(Y) = -\sum_i P(Y = a_i) \log P(Y = a_i)
\]

where \( a_i \) is the possible value of \( Y \). For continuous random variable vector, entropy is defined as:

\[
H(y) = -\int f(y) \log f(y) dy
\]

A Gaussian variable has the largest entropy among all random variables of equal variance. Negentropy \( J \) is thus defined as follows to measure the non-Gaussianity:

\[
J(y) = H(y_{gauss}) - H(y)
\]

However, the problem in using negentropy is the difficult in computation. Therefore, many ways of approximation have been proposed, some of which have very appealing properties such as robustness. Independent Components can be estimated by maximizing the approximation of negentropy.

**Minimizing mutual information** Finding the most non-Gaussian directions is more or less a heuristic way to estimate the independent components. Another approach, inspired by information theory, is minimization of mutual information. It is shown to end up with the same principle but gives a rigorous justification.

The mutual information between \( m \) random variables \( y_i, i = 1,2,\cdots, m \) is defined as:

\[
I(y_1, y_2, \cdots, y_m) = \sum_{i=1}^{m} H(y_i) - H(y)
\]

which is always non-negative, and zero if and only if the variables are statistically independent. Thus mutual information is a natural information-theoretic measure of the independence of random variables. It is shown that minimizing mutual information is roughly equivalent to finding maximally the non-Gaussian directions.

**Maximum likelihood** A very popular approach for estimating ICA model is maximum likelihood estimation. It is shown to be equivalent to minimizing mutual information.

For simplicity, assuming the mixing matrix \( A \) is square, ICA estimation is basically the estimation of its inverse \( W \), and ICA is estimated as:

\[
\hat{s}(t) = \hat{W}x(t)
\]

The log-likelihood takes the form:

\[
L = \sum_{i=1}^{T} \sum_{j=1}^{n} \log f_j(w_j x(t)) + T \log \det |W|
\]

where \( W = (w_1, w_2, \cdots, w_n)^T = A^{T} \), \( f_i \) is the density function of \( s_i \), \( T \) is the overall time.

**The informax principle** The informax principle was derived from a neural network viewpoint. It is based on maximizing the output entropy (or information flow) of a neural network with non-linear outputs.
Assume that $x(t)$ is the input to the neural network whose outputs are of the form $\phi_i(w_i^T x(t))$, where $\phi$ is some non-linear scalar function, and $w_i$ is the weight vector of the neurons. The entropy of the output:

$$L = H(\phi_1(w_1^T x), \phi_2(w_2^T x), \ldots, \phi_n(w_n^T x))$$

is maximized when $\phi$ is well chosen to estimate the ICA model. Indeed, several authors proved the surprising result that the principle of network entropy maximization, or “infomax”, is equivalent to maximum likelihood estimation when the non-linear $\phi$ used in the neural network is chosen as the cumulative distribution function corresponding to the densities $f_i$.

To sum up, the intuitive notion of maximum non-Gaussianity can be used to derive different objective functions (such as kurtosis) whose optimization enables the estimation of the ICA model. ICA can also be estimated from other viewpoints such as information theory and neural network. Surprisingly, these approaches are approximately equivalent.

### 3.2.3 Blind source separation with under-determined case

In the under-determined case, complete separation is usually impossible. This is easy to understand by considering a much simpler situation where the mixing matrix $A$ is known (recall that in BSS this does not hold), and there is no noise. Even then the set of linear equations (3.1) has an infinite number of solutions because there are more unknowns than equations, and the source vector $s(t)$ cannot be determined for arbitrary distributed sources.

However, separation may still be achievable in special instances at least. This topic was studied theoretically in [23]. The authors show that it is possible to separate the $n$ sources into $m$ disjoint groups if and only if $A$ has $m$ linearly independent column vectors, and the remaining $m-n$ column vectors satisfy the special condition that each of them is parallel to one of these $m$ column vectors.

**Sparse Representation** Without *a priori* knowledge, under-determined source separation is very difficult. Some additional assumptions have to be made. A very powerful assumption is that the sources have a parsimonious representation in a given basis, such as the time-frequency (T-F) representation. Some researchers have presented encouraging results. [24][25][30] If the source signals do not overlap in the time-frequency domain, high-quality reconstruction could be obtained. [26]

**Binary Masking, CASA, ICA** Good separation can also be obtained by applying a similar promising technique named binary time-frequency masking to the mixture. The technique is based on an assumption that spectrogram of the mixture is almost exactly the element-wise maximum of the original spectrograms, taken across a narrowband time-frequency plane. The approximation to spectrogram mixtures comes from a simple mathematical relationship:

$$\log(e_1, e_2) \approx \max(\log e_1, \log e_2) \quad \text{when } e_1 << e_2 \text{ or } e_2 << e_1$$

coupled with the sparsity of speech energy distribution in time-frequency. [26][27] For a binary mask, each T-F unit is either weighted by one or by zero. In order to reduce musical noise, smoother masks may also be applied. [28] An advantage of using a binary mask is that only a binary decision has to be made. Such a decision can be based on, e.g., clustering or direction-of-arrival. ICA has been used in different combinations with the binary mask. In CASA, the technique
of T-F masking has been commonly used for years, which has become popular in blind source separation. The combination of CASA and ICA is also a technique under investigation to solve the under-determined problem.

### 3.2.4 Blind source separation with correlated sources

The popular assumption of blind source separation is the independence of sources. When the independence assumption is slightly violated, ICA is still favorable. However, when the sources are highly correlated, those general techniques fail. It is very difficult to find a general solution since very limited *a priori* is available and how the sources are dependent is unknown.

[29] gives an algorithm which can roughly retrieve the original source signals even though they are not completely independent or not even uncorrelated (which is a considerably milder condition than independence). Thus, they can achieve in the BSS problem more than the current theory promises. However, as the author stated, the results still need more studies in the future.

[30] proposed a simple and efficient method for solving the linear instantaneous BSS problem with *n* sources and *m* observations. This approach is based on the Time–Frequency Version of Ratios Of Mixtures of source signals, and is therefore called “TIFROM”. It mainly relies on the assumption that a source is “visible”, i.e. that it occurs alone (as opposed to the other sources) in at least one tiny area in the TF plane. The proposed algorithm automatically determines such an area and then derives coefficients which e.g. allow one to cancel the contributions of this source from the observed signals. This makes it possible to separate all sources in various situations, e.g. when *n=* *m* and all sources are visible. This method is still of interest in other cases, e.g. for underdetermined mixtures: it then separates part of the sources. This approach requires *m* ≥ 2.

Several geometric-inspired algorithms can also separate sources that are correlated in a specific way. [31] provides a new of separating correlated images but it requires that a squared pdf with correlation "inside" the borders of the bounded pdf.

MSD ICA is also able to separate dependent sources. However, it assumes that a frequency band exists, where the sources are statistically independent. [32]

In summary, although there is not much literature dealing with correlated blind source separation problem, some algorithms exist, depending on the "kind of correlation".

### 3.3 Semi-Blind Source Separation

Compared with blind source separation, the semi-blind source separation requires partial information about the mixing process or sources. This additional informational is usually application dependent. Thus, semi-blind source separation algorithms are very ad hoc.
3.3.1 Semi-blind source separation with under-determined case

In BSS, under-determined problem is difficult to deal with. However, due to the additional available information in semi-blind problem, it is usually a little easier to find the solution.

A typical class of semi-blind source separation algorithms is one microphone source separation or single-channel source separation which is a hot research area recently. As an extreme case of under-determined problem, some additional constrains are required, which can be the limitations on the possible forms of source signals. In practice, real-world sound sources of interest have structured and constrained properties, such as stationarity, periodicity and limited spectral variation. The task of monaural source separation can be viewed as a problem of suitably capturing and applying these constrains.

Using training phase to obtain the strong prior knowledge

A training phase can be used to obtain strong prior knowledge about the source model. And later, in separation phase, the mixture is separated based on the knowledge.

One way of using this technique is to model the density function of filtered signals in training phase. The basis filters, such as DFT filters [33], DCT filters [34], adaptive filters learned by ICA [35] are chosen with specific properties such as independence of coefficients. After obtaining prior in the training phase, the separation is based on a Bayesian inference or maximum likelihood (usually used when Bayesian inference is not tractable).

Besides directly modeling the density function of filtered signals using basis filters, modeling could happen in feature space using vector quantization and clustering algorithms when proper features are extracted. Separation can then be achieved by finding the best-matching code-words with an observed mixture. This technique is especially useful for speech signal since many meaningful features are well known.

Normally, the source model is built up without phase information, which degrades the sound quality. A perceptually satisfactory model with phase information is also proposed recently [36].

3.3.2 Semi-blind source separation with correlated sources

Compared with under-determined cases, literature on source separation with correlated sources is even less. However, it does exist in some special applications.

[37] gives a typical example of separating correlated sources with a prior. The structure of source covariance matrices and the mixing matrix are known. The algorithm parameterizes the mixing matrix and source covariance matrices first. The second-order statistics is used to identify those parameters.
3.4 Monaural Speech Separation

Speech is a main class of acoustic signals. This section summarizes the monaural speech separation techniques, most of which, however, are common approaches to any acoustic signals.

In recent years, in order to combat real situations encountered with speech processing applications, the robustness of proposed techniques has been the main concern of researchers. In this light, single channel (or monaural) speech segregation problem has gained exclusive attention. [20] gives an overview of the monaural speech segregation techniques.

The monaural speech segregation techniques can be categorized into two broad classes: (i) source-driven approaches and (ii) model-driven approaches.

In source-driven methods, the signal is extracted without any \textit{a priori} knowledge. They can also be divided into two categories: computational auditory scene analysis (CASA) and blind source separation. The CASA method relies on psychoacoustic cue such as pitch, onset, offset and continuity, for separation. The main drawbacks of this method are the difficulty to mitigate crosstalk and the difficulty to separate out the unvoiced regions. The other subdivision blind source separation (BSS) has been studied for nearly two decades and can be traced back to 1986. If the requirements of BSS methods are satisfied, they would be more effective than CASA methods. However, the requirements for BSS are normally very high because of the blindness. One of the requirements for typical BSS is that the number if observations must be at least equal to the number of sources, a condition which is not held when only one microphone is available. Recently, some efforts have been made to extend BSS to the one microphone speech segregation problem with requirement of \textit{a priori} knowledge. The fusion of CASA and BSS is also under investigation today, see [21].

The second class of speech separation techniques is commonly referred to as the model-based or model-driven separation and completely relies on \textit{a priori} knowledge. The speech signal is modeled using well-known statistical speech modeling techniques such as Hidden Markov Models (HMM), Gaussian Mixture Model (GMM), or Vector Quantization (VQ). After the model is trained, the mixture signal is separated based on some criteria (e.g., minimum mean square error, likelihood ratio). Although the separated speech obtained using model-based techniques exhibits high quality, signal-dependency and computational complexity are two major disadvantages associated with this approach.

3.5 Summary

This chapter provides an overview of a great variety of source separation techniques. To clarify the overall picture of the problem, source separation problem is divided into two categories: blind source separation and semi-blind source separation. The difference between them lies in the amount of prior knowledge about the mixing process and source characteristics. In each category, over-determined, even determined and under-determined cases are studied one by one. Moreover, correlated source separation is also paid enough attention to. In general, under-determined and correlated source separation is difficult and strong \textit{a priori} is normally required. The algorithms are also very ad-hoc, which, however, are shown in the next chapter to be vital for feedback
cancellation systems in hearing aids. The last section of this chapter summarizes the monaural source separation techniques due to its relevance to feedback cancellation system.
Chapter 4 Source Separation for Feedback Cancellation

The previous two chapters review the techniques of feedback suppression and source separation, which appear to be distinct areas since they are developed from different starting points and in very different ways. However, this chapter will investigate the possibility of using source separation techniques for feedback cancellation. Very little literature has ever dealt with this problem. Two proposals are investigated and one novel approach is proposed in this chapter. Due to the immaturity of source separation techniques for difficult situations, such as in under-determined cases and with correlated sources, issues remain in the new proposal.

This chapter is organized as follows: Section 4.1 discusses the purpose of using source separation in feedback cancellation systems. Section 4.2 then deals with the possibility of using it based on the review in the last two chapters. Section 4.3 and 4.4 are two proposals belonging to indirect and direct ways of using source separation techniques respectively. In section 4.5, a novel source separation algorithm is proposed to deal with the difficult under-determined case with correlated sources. Two strategies are introduced to incorporate the new algorithm into feedback cancellers. The performance is evaluated in a standard feedback canceller and an advance feedback canceller invented by GN Resound. Section 4.6 gives a summary to the whole chapter.

4.1 Aim of using source separation in feedback cancellation system

As described in Chapter 2, there are a great number of techniques for feedback suppression in hearing aids, among which feedback cancellation system has been the focus due to its outstanding performance in most cases. Nowadays, people are especially concerned with several important problems regarding feedback cancellation system. The first problem is the converging speed of the adaptive filters. Feedback path is changing with the environment from time to time. Feedback canceller has to track the change sufficiently fast in order to remove the feedback signal. Thus, the converging speed is vital. The second problem is bias. All the feedback cancellers share a common problem of bias when the input signal is spectrally colored, such as tonal signal and music, which limits the performance of feedback canceller. The aim of using source separation techniques is to attack these two problems.

Among a variety of feedback cancellers, three of them are selected as possible candidates for the new proposals: the simplest canceller with a NLMS (normalized LMS) adaptation algorithm, a simple canceller with filtered-X modified LMS adaptation algorithm and an advanced feedback canceller developed by GN Resound.

4.2 Usage of source separation in feedback cancellation

There are two ways to use the techniques of source separation in feedback cancellation in order to enhance the performance: indirect and direct way.
Indirect usage of source separation techniques refers to the methods that take advantage of source separation ideas or principles, adopt new adaptation algorithms based on that, but do not separate the mixture signal explicitly. Direct usage, on the contrary, separates the mixture signal explicitly.

### 4.2.1 Indirect usage of source separation techniques in feedback cancellation

The way of indirect usage is basically comparing the feedback problem with some typical source separation models, making the same assumption and deriving similar adaptation schemes. The comparable source separation models are mainly from neural network, which has been one of the principle origins of source separation techniques, since the adaptive structure is similar to feedback canceller.

One approach of indirect usage of source separation techniques based on ICA was investigated in section 4.3.

### 4.2.2 Direct usage of source separation techniques in feedback cancellation

Direct usage of source separation techniques in feedback cancellation is more natural since feedback cancellation itself can be regarded as a process of source separation. The microphone signal $x(n)$ is a mixture of input signal $x(n)$ and feedback signal $f(n)$. Feedback canceller tries to estimate $f(n)$ by using adaptive algorithm and remove it from $s(n)$. Thus, in essence, it separates $s(n)$ into two parts. In addition, $e(n)$ can be regarded as a mixture of input signal and error signal too. Separating $e(n)$ may also help with the feedback canceller.

Suppose only one microphone is equipped with hearing aids, there will be only one observation available but with two sources to estimate. If there are two microphones, there are two observations but with four sources to estimate. For either of them, the problem is an under-determined source separation. Without loss of generality, this project deals with the one microphone case. Due to the short delay in the forward path of hearing aids, for spectral colored signals, such as tonal signal, voiced speech and music, the two sources, $f(n)$ and $x(n)$, are highly correlated. Therefore the source separation problem in feedback cancellation is an under-determined case with correlated sources, which is an extremely difficult problem.

The literature review of source separation techniques in Chapter 3 shows that existing algorithms deal with either under-determined case of correlated case, but none of them deals with both. As pointed out in Chapter 3, source separation problem with either under-determined case or correlated sources requires strong assumption or prior knowledge. It is reasonable to argue that, under-determined source separation problem with correlated sources will require stronger assumption or more prior knowledge.

The following table gives the possibility of using general source separation techniques for feedback cancellation based on the review in Chapter 3 to select possible candidates for direct usage of source separation techniques for feedback cancellation system.
4.2 Usage of source separation in feedback cancellation

<table>
<thead>
<tr>
<th>Source separation techniques</th>
<th>Possibility level</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICA, MSDICA and etc</td>
<td>low</td>
<td>ICA assumes the independence of sources. Moreover, it alone still has problem with under-determined cases.</td>
</tr>
<tr>
<td>Binary masking</td>
<td>low</td>
<td>The feedback signal and input signal are highly correlated, corresponding to a strong overlap in the frequency domain at any time. The strategy of binary masking is to assign spectrum at each frequency bin completely to either of the signal. Thus, it doesn’t help.</td>
</tr>
<tr>
<td>CASA</td>
<td>medium</td>
<td>Human beings are able to distinguish feedback signal and input signal when the feedback is not too strong. CASA is based on perceptual cues. If proper cues are used as features, it may help with separation of feedback signal and input signal. Moreover, CASA combined with BSS techniques forms a promising area.</td>
</tr>
<tr>
<td>TIFROM</td>
<td>low</td>
<td>TIFROM assumes that some areas exist in frequency-time domain where overlapping is very small. It is not a proper method for feedback cancellation since feedback signal coincides a lot with the input signal in the spectrum.</td>
</tr>
<tr>
<td>Training phase +ML/BI (Bayesian Inference)</td>
<td>high</td>
<td>This is by far the most promising approach since it has little assumption about the properties of sources, such sparsity, correlation of the sources and etc. In addition, adequate prior knowledge can be obtained in the training phase to deal with the difficult under-determined case.</td>
</tr>
</tbody>
</table>

When blind source separation techniques are considered to solve this under-determined source separation problem with correlated sources, assumptions about sources have to be made strongly. The mixing process is known as

\[ s(n) = x(n) + f(n) \] (4.1)
Since the signal hearing aids processed is mainly speech, widely accepted assumptions regarding speech can be made. An algorithm based on AR modeling of speech is thus investigated in section 4.4.

When semi-blind source separation techniques are taken into consideration, prior knowledge has be obtained before the separation. Inspired by methods in section 3.3.1 and 3.4, a new novel idea by using the existing feedback canceller to provide prior knowledge is proposed. The algorithm is elaborated in section 4.5.

4.3 ICA-based LMS: increase the rate of convergence

Advanced adaptive techniques, such as RLS (Recursive Least-Squares), sub-band adaptive filters and so on, can be used to increase the converging speed. The basic idea of ICA also provides a possible improvement to the converging speed of conventional adaptive algorithms.

[38] claims that the ICA-based adaptive algorithm is applicable to adaptive noise cancelling (ANC), the diagram of which is shown below, with much better performance than that of LMS algorithm.

![Figure 4-1 The diagram of adaptive noise cancelling problem](image)

Figure 4-1 shows the diagram of ANC problem in a special way that resembles CAF of hearing aids in Figure 2-3. $x(n)$ is referred to as primary input. Noise source $n_1(n)$ is called reference noise signal. It is noted that ANC and CAF are very similar if the convolutive channel is regarded as the feedback path. The only difference is that in CAF, there is a forward path, i.e. hearing-aid process, connecting the input to the feedback path $y(n)$ with $e(n)$, while in ANC, the noise source $n_1(n)$ is not connected with the unmixed signal $u(n)$.

In ANC, the independence between $u(n)$ and $n_1(n)$ is assumed. If $y(n)$ and $e(n)$ are also assumed to be independent, the techniques used in ANC also apply to CAF. Based on this assumption and [38], a new adaptive filter is proposed in [39]. The learning rule is derived by maximizing the entropy. In this system, the Jacobian can be expressed as:

$$J = \frac{\partial y_1}{\partial x} \frac{\partial y_2}{\partial n_1} - \frac{\partial y_1}{\partial n_1} \frac{\partial y_2}{\partial x} = \frac{\partial y_1}{\partial u} \frac{\partial y_2}{\partial v}$$

(4.2)
where \( y_1 \) and \( y_2 \) are outputs of nonlinear functions approximating the CDF (Cumulative distribution function) of the signal \( s(n) \) and the noise \( n_1(n) \) respectively. \( v \) is a dummy output and \( v = n_1 \). By maximizing \( \log|f| \), the learning rule is obtained as follows:

\[
\Delta w(k) = -\mu \frac{\partial \log|f|}{\partial w(k)} = -\mu \phi(u(j))n_1(j - k) \tag{4.3}
\]

where the score function \( \phi(u(j)) \) is

\[
\phi(u(j)) = \left( \frac{\partial y_1}{\partial u} \right)^{-1} \frac{\partial^2 y_1}{\partial u^2}
\tag{4.4}
\]

whereas in conventional LMS

\[
\Delta w(k) = -\mu u(j)n_1(j - k) \tag{4.5}
\]

The difference between the LMS algorithm and the ICA-based approach comes from the existence of the score function. Introducing nonlinearity to the LMS algorithm has been studied by many researchers to improve the properties. The reason for the improvement is that LMS only removes noise components of the primary input signal based on second-order statistics. However there may exist many other components in the primary input signal which depend on the noise reference signal in higher-order statistics.

When the probability density function of the output signal \( u(n) \) approximates Gaussian distribution, the score function is:

\[
\phi(u(j)) = -\frac{\partial}{\partial u(j)} \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{u(j)^2}{2\sigma^2}} \right) = -\frac{u(j)}{\sqrt{2\pi\sigma^2}} \tag{4.6}
\]

This is equivalent to NMLS algorithm.

When the probability density function of the output signal \( u(n) \) approximates the Laplacian distribution, the score function is:

\[
\phi(u(j)) = -\frac{\partial}{\partial u(j)} \left( \frac{1}{\sqrt{2}} e^{-\sqrt{2}|u(j)|} \right) = \sqrt{2} \text{sgn}(u(j)) \tag{4.7}
\]

The score function \( \text{sgn}(u(j)) \) is superior to \( u(j) \) in conventional LMS when the input is speech since speech signal normally follows Laplacian distribution. The LMS algorithm based on Eq. (4.7) is also referred to as Sign LMS.

Sign LMS is therefore embodied in CAF to see if it can help increase the converging speed of adaptive filter. The simulation is setup the same way as in Chapter 2. Responses of microphone, receiver, amplifier and the un-amplified acoustic feed-forward path are simply ignored. The hearing-aid process \( h(n) \) is a simple delay of 1000 samples, which are used to de-correlate \( y(n) \) and
\( e(n) \) as much as possible. The gain of \( h(n) \) is 0 dB. Both the real feedback path and the adaptive filter are 120-order. The code is shown in the appendix.

When the input signal is a Laplacian noise, by doing numerous experiments, the best performance for NLMS algorithm was found with the step size \( \mu = 0.005 \) and for Sign LMS with the step size \( \mu = 0.0005 \). The two methods are evaluated by calculating the misalignment (in dB) of impulse response (see section 2.6). The result is shown below:

![Figure 4-2 The compare between NLMS and Sign LMS when the input is Laplacian noise](image)

When the input signal is a male speech, the best performance for NLMS algorithm was found with the step size of 0.003 and for Sign LMS with the step size 0.0001. The result is shown below:

![Figure 4-3 The compare between NLMS and Sign LMS when the input is male speech](image)
For real speech, the convergence is not that stable for either NLMS or sign LMS due to the bias problem. Thus smaller step size has to be used which leads to slower convergence in Figure 4-3. From Figure 4-2 and Figure 4-3, it can be concluded that Sign LMS converges faster than NLMS and has smaller misalignment (1~2 dB less) for real speech. It is also noted that Sign LMS normally has smaller step size to achieve the best performance compared with NLMS.

However, 1000 samples delay in $h(n)$ to de-correlate $e(n)$ and $y(n)$ is not a practical setup since 5 ms delay deteriorates the sound perception and speech intelligibility [11]. For a hearing aid, the typical forward delay is around 1 ms~5 ms, i.e. 16~80 samples with 16 kHz sampling frequency. To model the realistic application, 3 ms delay was used. With this setup, the best performance for NLMS algorithm was found with the step size of 0.0005 and for Sign LMS with the step size 0.00003. The result is shown below:

![Figure 4-4 The compare between NLMS and Sign LMS when the input is male speech and the delay is very short](image)

Compared with Figure 4-3, the misalignment is much bigger due to the bias resulted from large cross-correlation when the delay in the forward path is very short. The performance of NLMS and Sign LMS is very similar. Thus in real application, except for the reduction of computation load, the benefit of Sign LMS is marginal and not very interesting.

### 4.4 Linear prediction: remove the bias

In the updating algorithm of adaptive filter in section 2.3, $e(n)$ is not a pure error signal but containing the input signal $x(n)$ inevitably. The bias of adaptive filter is thus resulted from the cross-correlation term $\Gamma_{w_1x_1}$. If $x(n)$ can be removed from $e(n)$, an unbiased adaptive scheme can be obtained. More practically speaking, if $x(n)$ can be approximated, the performance of feedback canceller should also be improved since the cross-correlation is reduced. This idea is shown below [40]:

In order to predict the input signal \( x(n) \), the Forward Linear Predictor (FLP) is utilized [40]. The estimation of \( e(n) \) is obtained by:

\[
\hat{e}(n) = \sum_{i=1}^{M} w_{LP}(i) * e(n-i)
\]

where \( w_{LP} \) is the prediction coefficients, \( M \) is the order. For speech modeling, the prediction order varies from 2 to 18 [40].

The remaining error is:

\[
e_i(n) = e(n) - \hat{e}(n)
\]

The updating of prediction coefficients is based on a recursive least square algorithm [10]:

\[
\text{Table 4-2 Summary of RLS algorithm for Linear Prediction}
\]

Initialize the algorithm by setting

\[
\hat{w}_{LP}(0) = \mathbf{0}
\]

\[
P(0) = \delta^{-1} \mathbf{I}
\]

and

\[
\delta = \begin{cases} 
\text{small positive constant for high SNR} \\
\text{large positive constant for low SNR}
\end{cases}
\]

For each instant of time, \( n = 1, 2, \ldots \), compute

\[
\pi(n-1) = P(n-1)e(n-1)
\]

\[
k(n-1) = \frac{\pi(n-1)}{\lambda + e^{H}(n-1)e(n-1)}
\]

\[
\xi(n) = e(n) - \hat{w}_{LP}(n-1)\hat{e}(n-1)
\]

\[
\hat{w}_{LP}(n) = \hat{w}_{LP}(n-1) + k(n-1)\xi(n)
\]

\[
P(n) = \lambda P(n-1) - \lambda P(n-1)k(n-1)e^{H}(n-1)P(n-1)
\]
where
\[
\mathbf{e}(n) = [e(n), e(n-1), \ldots, e(i - M + 1)]^T
\]
\[
\hat{\mathbf{w}}_{LP}(n) = [w_{LP,1}(n), w_{LP,2}(n), \ldots, w_{LP,M}(n)]^T
\]

Assume that the input signal \( x(n) \) dominates \( e(n) \), \( \delta \) is set as a small value 0.01. M is set to 10, a medium order, since too high an order will result in over-fitting of noise data. \( \lambda \) is a forgetting factor, depending on the characteristics of the signal to be predicted and satisfying \( 0 \ll \lambda < 1 \). For a chosen male speech, the optimal \( \lambda \) is found by simulations. The results are shown below (the code is shown in the appendix):

![Figure 4-6 Optimal forgetting factor for clean speech and noisy speech in linear prediction](image)

From the figures, it is seen that the optimal \( \lambda \)'s for clean speech and noisy speech with SNR=0 dB are 0.986 and 0.995 respectively. \( \lambda \) is thus chosen as 0.99, which is between the two extreme cases.

The same speech is input to a standard feedback canceller (modified filtered-X based LMS) tuned to the best performance as a reference to evaluate the algorithm. The performance is measured by ASG since filtered-X is used in the feedback canceller emphasizing the estimation on critical bands where feedback whistling usually occurs. The parameters are listed below:

| Test signals: | male speech (30 seconds, 16 kHz sampling rate) |
| Feedback canceller: | A standard feedback canceller with filtered-X modified LMS adaptation algorithm tuned to best performance |
| Feedback path: | Measured in real life, static. |

| Forward path of hearing aids: | A delay (8 samples, i.e., 0.5 ms) + Gain (21 dB) for |
Performance measure of feedback canceller:
Additional stable gain

Adaptive FIR filter:
16 orders, \( \mu = 1e^{-6}, \lambda = 0.995, c = 1e^{-5} \)

High pass filter (filtered-X): [0.6667 -0.3333]

DC filter: [1 -1]

Optimal delay in the adaptive path is found by an algorithm similar to initialization for GN Resound’s feedback canceller to give a best match between feedback path and adaptive path. The optimal step size of LMS \( \mu = 1e^{-6} \) is found by numerous simulations.

Linear predictor parameters:
\( \delta = 0.01, M = 10, \lambda = 0.99 \)

The result is shown below:

![Graph](image_url)

Figure 4-7 Performance compare between standard feedback canceller with and without linear prediction

The performance is not as good as expected and even worse than the standard feedback canceller. The reason can be explained by the following figures:
4.5 Weighted adaptive Wiener filter: A complicated scheme

In Figure 4-8, it is seen that the estimation of input signal $x(n)$ is very accurate, indicating that linear predictor works fine. However, the aim is to use $e_1(n)$ to model the error of estimation of feedback signal from feedback canceller, i.e.:

$$e(n) - \hat{e}(n) \approx f(n) - v(n) = e(n) - x(n)$$  \hspace{1cm} (4.10)

This fails though $x(n)$ is very well estimated by $\hat{e}(n)$. The reason is that $x(n)$ is much bigger than $f(n)$, the relatively small estimation error of $x(n)$ is very large for $f(n)$.

Although the performance is not exciting, the idea is still worth investigating. A more advanced algorithm with similar idea from system identification point of view to remove the bias is recently proposed in [9]. Due to the limit of time in this project, the further investigation of this idea is skipped.

4.5 Weighted adaptive Wiener filter: A complicated scheme

The new algorithm proposed in this section is inspired by source separation techniques described in section 3.3.1 and 3.4. The source separation algorithm itself and the way to help feedback canceller are novel.
4.5.1 General ideas of the new algorithm

As described in section 4.2.2, under-determined source separation problems with correlated sources are so difficult to address that no mature and robust algorithms have been proposed. Thus, though feedback cancellation itself can be regarded as a source separation problem, due to the difficulty in solving the problem satisfactorily, source separation techniques should only be used to help the feedback canceller instead of replacing it.

Table 4-1 shows that the semi-blind source separation technique utilizing training phase to obtain prior knowledge is especially promising for feedback cancellation problem due to its little assumption about sources and strong prior knowledge for separation phase. However, the question is where this kind of information should be obtained.

In section 4.4, the speech characteristics were actually utilized a little as assumption on the source structure. In this proposal, instead of making assumptions about source, prior information is extracted in the training phase. However, in hearing aids, off-line training, e.g. in [33], is not possible since feedback signal is always present and mixed with input signal. Even if an additional microphone is used to separate feedback signal and input signal physically in the training phase, when the characteristics of feedback path or input signal change, the prior information is not valid any more. Therefore, the training phase in the proposal is on-line. An on-line training has to be performed.

The aim of the source separation is to separate $x(n)$ and $f(n)$ from $s(n)$. Since the training is on-line, they are not available in the training phase. By assuming that the feedback canceller is working well at the moment, which means:

$$v(n) \cong f(n) \quad e(n) \cong x(n)$$

(4.11)

the source separation module does the training based on $e(n)$ and $v(n)$ instead. It models the joint probability density function of $e(n)$ and $v(n)$. After the training phase, Bayesian inference is utilized to unmix $s(n)$ optimally in the statistical sense. and output the result to feedback canceller. The proposal is depicted in the following diagram:

![Figure 4-9 The diagram of feedback cancellation with source separation](image-url)
4.5 Weighted adaptive Wiener filter: A complicated scheme

It is noted that Figure 4-5 is a special case of Figure 4-9. \( ss(n) \) is the information source separation provides to help the feedback canceller.

The training and separation can be performed in time domain or any domain spanned by proper basis functions, such as DFT (Discrete Fourier Transform) domain. In different frequency bins, since the feedback path has different gains and phase shifts, the feedback signal and input signal must be correlated differently. Therefore time-domain density modeling does not have enough resolution to describe the difference between frequency bins. Usually, as proposed in [33], the probability density functions of the two sources are modeled in DFT domain, where most hearing-aid algorithms also work. The proposed source separation algorithm is thus chosen to perform also in DFT domain to take these advantages. Another advantage of the DFT domain is that the spectral representation may be often simpler than the time representation and frequency components may be far less correlated than time samples. Other possible domains will also be investigated in the section 4.5.6.

The following section is organized as follows: section 4.5.2 and 4.5.3 prepares the basic knowledge and deals with the training phase. Section 4.5.4 derives the separation. Section 4.5.5 provides the simulation results showing how the proposed algorithm works. Section 4.5.6 discusses the algorithm in DCT domain. Section 4.5.7 proposes several ways of embodying the information obtained from source separation in the feedback canceller. Lastly, section 4.5.8 gives the simulation results.

4.5.2 Complex Gaussian distribution and Gaussian statistical model

The signals processed in the project are not deterministic but random. Since source separation algorithm is carried out in DFT domain which is complex with magnitude and phase, the complex distribution and random process in DFT domain have to be learned beforehand. These topics are rarely covered in normal random process literature. Thus this section tries to give brief introduction, Some of the formulas in this section are derived especially for this project.

Complex Gaussian distribution A complex random variable \( \tilde{x} = u + jv \), where \( \sim \) indicates that the variable is complex, follows the Gaussian distribution if its real part \( u \) and imaginary part \( v \) are independent Gaussian variables with the same variance, denoted as \( \sigma^2 \). The probability density function of \( \tilde{x} \) is [41]:

\[
p(\tilde{x}) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{1}{\sigma^2} |\tilde{x} - \tilde{\mu}|^2 \right)
\]

(4.12)

where \( \tilde{\mu} = E[\tilde{x}] = \mu_u + j\mu_v \). Or it can also be written as \( \tilde{x} \sim CN(\tilde{\mu}, \sigma^2) \).

When the real and imaginary part of a random complex variable are independent and have the same variance, it is called to be circular symmetric. This is mandated for a complex Gaussian variable because otherwise the concise form (4.12) doesn’t apply and some desirable properties do not hold.

Now we consider the multivariate complex Gaussian distribution. Suppose there are two complex Gaussian variables \( \tilde{x} = u_1 + jv_1 \), \( \tilde{x} \sim CN(\tilde{\mu}_1, \sigma_1^2) \) and \( \tilde{y} = u_2 + jv_2 \), \( \tilde{y} \sim CN(\tilde{\mu}_2, \sigma_2^2) \) that follow bivariate Gaussian distribution, i.e.
\( \tilde{z} \sim CN(\mu_z, C_z) \) \hspace{1cm} (4.13)

where \( \tilde{z} = [\tilde{x}, \tilde{y}]^T \), \( \tilde{\mu}_z = [\tilde{\mu}_x, \tilde{\mu}_y]^T \) and \( C_z \) is the covariance matrix.

The distribution of random vector \( \tilde{z} \) depends on \( \tilde{x} \) and \( \tilde{y} \) in a way that the Gaussian probability density function is not expressible only in terms of \( \tilde{z} \) \([42]\). \([43]\) introduced a special case, restricting the covariance matrix of \( \tilde{x} \) and \( \tilde{y} \), that generalizes all the basic statistical theory related with the real Gaussian distribution. In brief, provided the real and imaginary components obey the covariance relations

\[
\text{Cov}(u_1, u_2) = \text{Cov}(v_1, v_2) \quad \text{Cov}(u_1, v_2) = -\text{Cov}(u_2, v_1) \quad (4.14)
\]

the joint distribution has the simple form

\[
p(\tilde{z}) = \frac{1}{\pi^2|C_z|} \exp \left[ -((\tilde{z} - \tilde{\mu}_z)^H C_z^{-1}(\tilde{z} - \tilde{\mu}_z) \right] \quad (4.15)
\]

where

\[
C_z = \begin{bmatrix}
\text{Cov}(\tilde{x}, \tilde{x}) & \text{Cov}(\tilde{x}, \tilde{y}) \\
\text{Cov}(\tilde{y}, \tilde{x}) & \text{Cov}(\tilde{y}, \tilde{y})
\end{bmatrix}
\]

As \( \tilde{x} \) and \( \tilde{y} \) are both complex Gaussian variables complying with (4.14), the covariance matrix can be calculated as:

\[
\text{Cov}(\tilde{x}, \tilde{x}) = E[(\tilde{x} - \tilde{\mu}_x)^*(\tilde{x} - \tilde{\mu}_x)] = \sigma_1^2
\]

\[
\text{Cov}(\tilde{y}, \tilde{y}) = E[(\tilde{y} - \tilde{\mu}_y)^*(\tilde{y} - \tilde{\mu}_y)] = \sigma_2^2
\]

\[
\text{Cov}(\tilde{x}, \tilde{y}) = E[(\tilde{x} - \tilde{\mu}_x)^*(\tilde{y} - \tilde{\mu}_y)] = 2\text{Cov}(u_1, u_2) + 2j\text{Cov}(u_1, v_2) = \sigma_{12}
\]

\[
\text{Cov}(\tilde{y}, \tilde{x}) = E[(\tilde{y} - \tilde{\mu}_y)^*(\tilde{x} - \tilde{\mu}_x)] = 2\text{Cov}(u_1, u_2) - 2j\text{Cov}(u_1, v_2) = \sigma_{12}^* \quad (4.17)
\]

Thus the covariance matrix \( C_z \) is:

\[
C_z = \begin{bmatrix}
\sigma_1^2 & \sigma_{12} \\
\sigma_{12}^* & \sigma_2^2
\end{bmatrix}
\]

Assume that \( \tilde{x} \) and \( \tilde{y} \) both have zero mean, (4.15) can be expanded as follows:

\[
p\left( \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} \right) = \frac{1}{\pi^2(\sigma_1^2\sigma_2^2 - |\sigma_{12}|^2)} \exp \left[ - \frac{1}{(\sigma_1^2\sigma_2^2 - |\sigma_{12}|^2)} (\sigma_1^2|\tilde{x}|^2 + \sigma_2^2|\tilde{y}|^2 - 2\text{Re}(\sigma_{12}^*\tilde{x}^*\tilde{y}) \right] \quad (4.19)
\]

Now consider the sum of \( \tilde{x} \) and \( \tilde{y} \):

\[
\tilde{w} = \tilde{x} + \tilde{y} = (u_1 + u_2) + j(v_1 + v_2) \quad (4.20)
\]

Assume that \( u_1 \) and \( u_2 \) follow 2-D Gaussian distribution and so do \( v_1 \) and \( v_2 \). \( \tilde{w} \) is not a complex Gaussian variable anymore since the real and imaginary parts are not independent any more.

However, the distribution could be described by real joint Gaussian distribution since the complex joint Gaussian distribution may also be represented by the 4-variate real Gaussian vector \( [u_1, u_2, v_1, v_2]^T \) with the covariance matrix:
4.5 Weighted adaptive Wiener filter: A complicated scheme

\[ C = \begin{bmatrix} \frac{\sigma_1^2}{2} & \text{Cov}(u_1, u_2) & 0 & \text{Cov}(u_1, v_2) \\ \text{Cov}(u_1, u_2) & \frac{\sigma_2^2}{2} & -\text{Cov}(u_1, v_2) & 0 \\ 0 & -\text{Cov}(u_1, v_2) & \frac{\sigma_1^2}{2} & \text{Cov}(u_1, u_2) \\ \text{Cov}(u_1, v_2) & 0 & \text{Cov}(u_1, u_2) & \frac{\sigma_2^2}{2} \end{bmatrix} \quad (4.21) \]

Now the complex random variable \( \tilde{w} \) is regarded as a 2-D vector:

\[ \tilde{w} = \begin{bmatrix} u_1 + u_2 \\ v_1 + v_2 \end{bmatrix} \quad (4.22) \]

Since \([u_1, u_2, v_1, v_2]^T\) is a 4-variate real Gaussian vector, \( w \) is also a Gaussian vector. The covariance matrix can be calculated as follows by referring to (4.14):

\[ \text{Cov}(u_1 + u_2, u_1 + u_2) = \frac{\sigma_1^2}{2} + \frac{\sigma_2^2}{2} + 2\text{Cov}(u_1, u_2) \]
\[ \text{Cov}(v_1 + v_2, v_1 + v_2) = \frac{\sigma_1^2}{2} + \frac{\sigma_2^2}{2} + 2\text{Cov}(v_1, v_2) \]
\[ \text{Cov}(u_1 + u_2, v_1 + v_2) = \text{Cov}(v_1 + v_2, u_1 + u_2) = 0 \quad (4.23) \]

Thus when assuming \( \tilde{x} \) and \( \tilde{y} \) both have zero mean, Gaussian vector \( w \) follows \( w \sim N(\theta, C) \), where

\[ C = \begin{bmatrix} \frac{\sigma_1^2}{2} + \frac{\sigma_2^2}{2} + 2\text{Cov}(u_1, u_2) & 0 \\ 0 & \frac{\sigma_1^2}{2} + \frac{\sigma_2^2}{2} + 2\text{Cov}(u_1, u_2) \end{bmatrix} \quad (4.24) \]

**Gaussian statistical model of DFT coefficients**  
After Fourier transformation all the values of the frequency components at a specific frequency can be collected. These are samples of a random variable whose distribution can also be computed.

A general conclusion about random process in the frequency domain is shown below [46]:

<table>
<thead>
<tr>
<th>Non-stationary processes</th>
<th>Correlations between frequency components, non-zero means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stationary processes</td>
<td>No correlations between frequency components, zero means, dependences and non-Gaussianity</td>
</tr>
<tr>
<td>Gaussian processes</td>
<td>Independence and Gaussianity in the frequency domain</td>
</tr>
</tbody>
</table>

However, since DFT is a Fourier transform for a periodical of signal and on discrete frequencies, some of the properties listed in the table above may not hold any more. The properties of DFT coefficients of a random process have to be found out since they will be used in the derivation of the algorithm.
Suppose \( x(n) \) is a zero-mean white Gaussian noise (WGN). It is proved in [10] that

\[
\begin{bmatrix}
\tilde{X}(m,1) \\
\vdots \\
\tilde{X}(m,\frac{L}{2}-1)
\end{bmatrix} \sim CN(0, N\sigma^2 I) \tag{4.25}
\]

where

\[
\tilde{X}(m, k) = \sum_{n=0}^{L-1} x((m-1)(L-M) + n + 1)e^{-\frac{2\pi i}{L}kn} \quad k = 0, 1, \ldots, L - 1 \quad m = 1, 2, \ldots
\]

\( L \) is the length of frame
\( M \) is the length of overlapping
\( m \) is the frame index

\( \tilde{X}(m,0) \) and \( \tilde{X}(m, \frac{L}{2}) \) are not interesting since they are DC and Nyquist components respectively.

\( \tilde{X}(m, k) \) \( k = \frac{L}{2} + 1, \ldots, L - 1 \) are also ignored due to the symmetry of DFT coefficients.

This conclusion shows that the DFT coefficients of \( x(n) \) are i.i.d. (independent identical distributed); the real and imaginary parts in each frequency bin are independent and have the same variance. The independence of frequency components exactly reflects its lack of structure.

However, for a general stationary process, the conclusion is not so strong but as follows [47]:

- The joint distribution of any finite set of elements belonging to the DFT of a block of length \( L \) from a stationary sequence converges to Gaussian as \( L \) approaches infinity if the elements of the sequence are strongly mixing (i.e., far separated elements are weakly dependent) and obey the (Lyapunov) condition that for any \( \delta > 0 \), the \( 2 + \delta \) th moment is finite.

- The joint distribution of any finite set of elements belonging to the DFT of a block of length \( N \) from a sequence of independent identically distributed random variables of finite variance converges to Gaussian as \( N \) approaches infinity.

A widely accepted assumption for stationary audio signal is that the DFT coefficients follow joint complex Gaussian distribution [48], i.e.,

\[
\begin{bmatrix}
\tilde{X}(m,1) \\
\vdots \\
\tilde{X}(m,\frac{L}{2}-1)
\end{bmatrix} \sim CN(0, N(diag(\sigma^2(k)))) \quad k = 1, 2, \ldots, \frac{L}{2} - 1 \tag{4.27}
\]

which means:

- Real and imaginary parts of \( \tilde{X}(m, k) \) are independent, Gaussian distributed and have the same variance.
- \( \tilde{X}(m, k_i) \) and \( \tilde{X}(m, k_z) \) are independent when \( k_i \neq k_z \)

For two correlated stationary audio signals \( x(n) \) and \( y(n) \), we would like to make the following assumptions to derive the algorithm in section 4.5.3 and 4.5.4:
4.5 Weighted adaptive Wiener filter: A complicated scheme

\[
\begin{bmatrix}
\tilde{X}_1(m,1), \cdots, \tilde{X}_i(m, \frac{L}{2}-1)
\end{bmatrix}^T \sim \text{CN}(0, N(diag(\sigma_i^2(k)))) \quad k = 1,2, \cdots, \frac{L}{2}-1 \quad i = 1,2
\]

\[
[\tilde{X}_1(m,k), \tilde{X}_2(m,k_2)]^T \sim \text{CN}(0, N\begin{bmatrix}
\sigma_1^2(k) & 0 \\
0 & \sigma_2^2(k_2)
\end{bmatrix}) \quad k_1 \neq k_2
\]

\[
[\tilde{X}_1(m,k), \tilde{X}_2(m,k)]^T \sim \text{CN}(0, N\begin{bmatrix}
\sigma_1^2(k) & \sigma_{12}(k) \\
\sigma_{12}(k) & \sigma_2^2(k)
\end{bmatrix})
\]

That is to say, the DFT coefficients of the two stationary processes are only correlated within the same frequency bin.

**Validation of assumptions** The assumptions made in (4.27) and (4.28) are not technically true since not all the stationary processes can fulfill the ideal model in the conclusions in [47]. Compared with the un-correlation properties of KLT (Karhunen L"{o}ve Transform) coefficients, the existence of correlation between DFT coefficients is also well known. Therefore, simulations are carried out to see if the assumptions we make hold approximately for typical signals processed in this project.

**Simulation 1:**
The AR process is a general type of stationary process when the roots of its characteristic equation are within the unit circle. Moreover, the speech can be modeled with a 2\textsuperscript{nd}~18\textsuperscript{th} order AR process [40]. Thus AR process is a representative test signal to investigate the assumptions in (4.27). In the simulation, the AR process is 10\textsuperscript{th} order with a set of arbitrary stationary coefficients. DFT is carried out for every frame. The DFT length and the frame length are both 256 samples. There is no overlapping between frames. 10000 frames are used to model the probability density function of spectrum accurately. The results are shown in Figure 4-10. It is seen that the assumption nearly holds except that in several frequency bins, the real and imaginary parts of the spectrum are not uncorrelated/independent. (“Uncorrelated” is equivalent to “independent” for a Gaussian random variable.)
Simulation 2:
To validate the assumptions made in (4.28), two stationary processes are generated: the first signal is the AR process used in simulation 1, denoted as random process X; the second random process Y is obtained by filtering X by the impulse response of a measured feedback path. The length of DFT, the length of frame, and the number of frames are the same as used in simulation 1. The results are shown in Figure 4-11. It is noted that the DFT coefficients of process X and Y almost follow complex Gaussian distributions. However, the imaginary and real parts are correlated in some frequency bins. For process Y, compared with process X in Figure 4-10, the DFT coefficients in some adjacent frequency bins are more correlated. It is also seen that although the two processes X, Y are correlated, their DFT coefficients only correlate in the same frequency bin.
4.5 Weighted adaptive Wiener filter: A complicated scheme

Simulation 3:
The setup is the same as simulation 2 except that the coefficients of AR process X are randomized. Numerous simulations are made to see if any severe violation of assumptions can be found. The result is that in very few cases the assumptions are violated severely and most of the simulations are similar to Figure 4-11. The worst case with largest violation is shown in Figure 4-12. It is seen that although all the other assumptions are severely violated, the assumption that DFT coefficients of process X and Y only correlate in the same frequency bin still holds.
Figure 4-12 An example of severe violation of assumptions for two stationary processes

Conclusions from simulation:
In most cases, the assumptions in (4.27) and (4.28) approximately hold. In some frequency bins, the real and imaginary parts of DFT coefficients are slightly correlated. In some adjacent frequency bins, the DFT coefficients are correlated. The assumption that DFT coefficients of process X and Y only correlate in the same frequency bin holds very well.

4.5.3 Training phase: GMM and EM algorithm

Generally speaking, there are three ways to model the probability density function: non-parametric, parametric and semi-parametric methods. The first method is to find a general form of density function. The typical example is histogram. It is unbiased but suffers from the problem called “the curse of dimensionality”. In our case, for each channel namely each frequency bin, the dimensionality is four since the discrete Fourier transform of $e(n)$ and $v(n)$ are complex, which is so high that it requires large amount of data for the training. The second method is to find the specific form of density function. The parameters are determined by maximum likelihood or Bayesian Inference. This method can evaluate new values rapidly. However, since the form is fixed, it cannot deal with situations in our case because different audio signals have very different form of probability density functions. The semi-parametric methods combine the merits of non-parametric and parametric methods. It can find a general form of density function, and the size of model only grows with the complexity of the problem not the training data points. A typical example of the semi-parametric methods is Gaussian Mixture Model (GMM).
The microphone signal \( s(n) \) is a mixture of input signal \( x(n) \) and feedback signal \( f(n) \), i.e.,
\[
s(n) = x(n) + f(n)
\] (4.29)

For the notation convenience, \( x(n) \) and \( f(n) \) will be denoted as \( s_1(n) \) and \( s_2(n) \) respectively. When over-lap-add DFT filter is used, (4.29) can be written as
\[
\tilde{S}(m,k) = \tilde{S}_1(m,k) + \tilde{S}_2(m,k)
\] (4.30)

where
\[
\begin{align*}
\tilde{S}(m,k) &= \sum_{n=0}^{L-1} s((m-1)(L-M) + n + 1)h(n)e^{-j\frac{2\pi kn}{L}} \quad k = 0,1,\cdots, L-1 \quad m = 1,2,\cdots \\
\tilde{S}_i(m,k) &= \sum_{n=0}^{L-1} s_i((m-1)(L-M) + n + 1)h(n)e^{-j\frac{2\pi kn}{L}} \quad i = 1,2
\end{align*}
\] (4.31)

\( L \) is the length of frame
\( M \) is the length of overlapping
\( h(n) \) is the window
\( m \) is the frame index

**Why and how to use GMM to model the density function of spectrum**

In this project, the joint distribution of \( \tilde{S}_1(m,k) \) and \( \tilde{S}_2(m,k) \) is modeled by GMM. There are two reasons for this choice: (I) GMM is a semi-parametric way of modeling the probability density function which combines the merits of parametric and non-parametric ways; (II) A more important reason is that usually audio signals are only locally stationary and contain various types of timbers and pitches. Therefore, their power spectral densities (PSD), which gives the variance as a function of frequency and could be regarded as a variance decomposition of the total energy of the process, is time-varying and has different shapes. The use of GMM takes into account the diverse structure of sounds through multiple PSD. Thus it is possible to model non stationary, still locally stationary signals. In fact, this model is quite general and leads to an adaptive Wiener filter scheme in the separation phase. [44]

There are two ways of modeling the joint distribution of \( \tilde{S}_1(m,k) \) and \( \tilde{S}_2(m,k) \) by GMM. In the first way the instant spectrum is regarded as a random vector: \( \left[ \tilde{S}_1(m,1), \cdots, \tilde{S}_1(m, \frac{L}{2} - 1) \right]^T \) \( i = 1,2 \). The joint distribution is
\[
p\left[ \left[ \tilde{S}_1(m,1), \cdots, \tilde{S}_1(m, \frac{L}{2} - 1) \right]^T, \left[ \tilde{S}_2(m,1), \cdots, \tilde{S}_2(m, \frac{L}{2} - 1) \right]^T \right).
\]

Due to the assumption that \( s_1(n) \) and \( s_2(n) \) only correlate in the same frequency bins, the covariance matrix is sparse. For each Gaussian component, the covariance matrix can be written as:
Chapter 4. Source Separation for Feedback Cancellation

\[
\begin{bmatrix}
\sigma_1^2(1) & 0 & \cdots & \cdots & 0 & \sigma_{12}(1) & 0 & \cdots & \cdots & 0 \\
0 & \sigma_1^2(2) & 0 & \cdots & \cdots & 0 & \sigma_{12}(2) & 0 & \cdots & \cdots \\
\vdots & 0 & \sigma_1^2(3) & 0 & \cdots & \cdots & 0 & \sigma_{12}(3) & 0 & \cdots \\
\vdots & \vdots & 0 & \ddots & \vdots & \cdots & 0 & \vdots & \ddots & 0 \\
0 & \cdots & \cdots & 0 & \sigma_1^2(L/2-1) & 0 & \cdots & \cdots & 0 & \sigma_{12}(L/2-1) \\
\sigma_{12}^*(1) & 0 & \cdots & \cdots & 0 & \sigma_2^2(1) & 0 & \cdots & \cdots & 0 \\
0 & \sigma_{12}^*(2) & 0 & \cdots & \cdots & 0 & \sigma_2^2(2) & 0 & \cdots & \cdots \\
\vdots & 0 & \sigma_{12}^*(3) & 0 & \cdots & \cdots & 0 & \sigma_2^2(3) & 0 & \cdots \\
\vdots & \vdots & 0 & \ddots & \vdots & \cdots & 0 & \vdots & \ddots & 0 \\
0 & \cdots & \cdots & 0 & \sigma_{12}^*(L/2-1) & 0 & \cdots & \cdots & 0 & \sigma_2^2(L/2-1)
\end{bmatrix}
\]

(4.32)

The second way uses GMM in each frequency bin, i.e., model \(p(\tilde{S}_1(m,k), \tilde{S}_2(m,k))\) for every \(k\). In this way the GMM of spectral of two correlated locally stationary signals is formulated as:

\[
p(\tilde{S}_1(m,k), \tilde{S}_2(m,k)) = \sum_{i=1}^{Q} \omega_i(k) p_G(\tilde{S}_1(m,k), \tilde{S}_2(m,k), \mathbf{C}_i(k))
\]

(4.33)

where \(Q\) is the number of components in GMM, \(k\) is the index to the \(k\)th frequency bin, \(i\) is the index to the \(i\)th Gaussian component, \(\omega_i(k)\) is the weight for the \(i\)th Gaussian component in the \(k\)th frequency bin, \(\sum_{i=1}^{Q} \omega_i(k) = 1\), \(p_G(\tilde{S}_1(m,k), \tilde{S}_2(m,k), \mathbf{C}_i(k))\) is the centered complex Gaussian distribution with covariance matrix \(\mathbf{C}_i(k)\), which is can be formulated as:

\[
\mathbf{C}_i(k) = \begin{bmatrix}
\sigma_{11}(k) & \sigma_{12}(k) \\
\sigma_{12}(k)^* & \sigma_{22}(k)
\end{bmatrix}
\]

\[
|\mathbf{C}_i(k)| = \sigma_{11}(k)\sigma_{22}(k) - |\sigma_{12}(k)|^2
\]

(4.34)

Therefore, by referring to (4.15) and (4.19), \(p_G\) can be written as:

\[
p_G(\tilde{S}_1(m,k), \tilde{S}_2(m,k), \mathbf{C}_i(k)) = \frac{1}{\pi^2 |\mathbf{C}_i(k)|} \exp\left(- \frac{1}{|\mathbf{C}_i(k)|} [\tilde{S}_1(m,k)]^\dagger \mathbf{C}_i(k) [\tilde{S}_1(m,k)] \right)
\]

\[
= \frac{1}{\pi^2 |\mathbf{C}_i(k)|} \exp\left(- \frac{1}{|\mathbf{C}_i(k)|} \left( |\tilde{S}_1(m,k)|^2 + |\tilde{S}_2(m,k)|^2 + 2 \text{Re}(\sigma_{12}(k)^* \tilde{S}_1(m,k) \tilde{S}_2(m,k)) \right) \right)
\]

(4.35)

The two ways are almost identical. The second is more straight-forward and easy to use. Thus, in this project, GMM is only applied in each frequency bin. (The slight difference between the two models is discussed in section 5.2.) The GMM is not necessarily used over all the frequencies. It can be used only in those critical frequency bands where feedback whistling usually occurs. In this way, the computation load can be reduced a lot.

**Estimation of parameters in GMM** In each frequency bin, by fixing the mean of each component as zero, there are totally \(4Q\) parameters to estimate including \(3Q\) in the covariance matrix \(\mathbf{C}_i\) and \(Q\) in the weights \(\omega_{k,j}\). A Bayesian learning algorithm EM (expectation maximization)
4.5 Weighted adaptive Wiener filter: A complicated scheme

is used to estimate those parameters. The standard procedure is to collect all the data and do iterative updates based on the following formulas [45] until some termination conditions, such as the maximum iteration number, are met. The updating formulas are:

\[
C_{j}^\text{new}(k) = \frac{\sum_{i=1}^{m} p_{k}^\text{old}(j \|[\tilde{S}_{1}(i,k), \tilde{S}_{2}(i,k)]^T) [\tilde{S}_{1}(i,k), \tilde{S}_{2}(i,k)]^H [\tilde{S}_{1}(i,k), \tilde{S}_{2}(i,k)]}{\sum_{i=1}^{m} p_{k}^\text{old}(j \|[\tilde{S}_{1}(i,k), \tilde{S}_{2}(i,k)]^T)}
\]

\[
p_{k}^\text{new}(j) = \omega_{j}^\text{new}(k) = \frac{1}{m} \sum_{i=1}^{m} p_{k}^\text{old}(j \|[\tilde{S}_{1}(i,k), \tilde{S}_{2}(i,k)]^T)
\]

\[
p_{k}^\text{new}(j \|[\tilde{S}_{1}(i,k), \tilde{S}_{2}(i,k)]^T) = \frac{p_{k}^\text{new}(j \|[\tilde{S}_{1}(i,k), \tilde{S}_{2}(i,k)]^T) p_{k}^\text{new}(j)}{\sum_{j=1}^{Q} p_{k}^\text{new}(j \|[\tilde{S}_{1}(i,k), \tilde{S}_{2}(i,k)]^T) p_{k}^\text{new}(j)}
\]

\[
= p_{G}(\begin{bmatrix} \tilde{S}_{1}(i,k) \\ \tilde{S}_{2}(i,k) \end{bmatrix}, C_{j}^\text{new}(k)) \omega_{j}^\text{new}(k)
\]

\[
= \frac{1}{\sum_{j=1}^{Q} p_{G}(\begin{bmatrix} \tilde{S}_{1}(i,k) \\ \tilde{S}_{2}(i,k) \end{bmatrix}, C_{j}^\text{new}(k)) \omega_{j}^\text{new}(k)}
\]

where \( p_{k}(j \|[\tilde{S}_{1}(i,k), \tilde{S}_{2}(i,k)]^T) \) is the probability that current \([\tilde{S}_{1}(i,k), \tilde{S}_{2}(i,k)]^T\) is from Gaussian component \(j\).

\( m \) is the number of samples

\( \text{old} \) refers to the estimation of last iteration

\( \text{new} \) refers to the estimation of current iteration

\( k = 1, 2, \ldots, \frac{L}{2} - 1; j = 1, 2, \ldots, Q \)

The EM algorithm is very sensitive to the initialization. In this project, it was found that a stable initialization method is to do the K-means clustering first, label the data and calculate the initial covariance and weights.

**On-line EM algorithm** The EM algorithm is an off-line algorithm, gathering all the data beforehand and doing iterative estimation afterwards. This is not a desired way for on-line algorithm. An alternative way is to store all the data coming so far and do the EM estimation according to formulas (4.36) every time when a new sample is coming. But this is an extremely computation demanding way. Thus, an easy on-line EM algorithm has to be derived. The on-line EM algorithm is useful when the statistics of the signal change over time.

Assume that the EM estimation has already converged to the maximum likelihood solution \( p_{k,m}(j), C_{j,m}(k) \) and \( p_{k,m}(j \|[\tilde{S}_{1}(i,k), \tilde{S}_{2}(i,k)]^T) i = 1, 2, \ldots, m \), where \( m \) denotes the solution is based on \( m \) samples. For every new sample, the estimation just needs to be updated a little instead of doing overall estimation again. To achieve this, the following approximation is made:
Chapter 4. Source Separation for Feedback Cancellation

\[ \sum_{i=1}^{m+1} p_{k,m+1}(j|\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T) \cong \sum_{i=1}^{m+1} p_{k,m}(j|\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T) \]

\[ = \sum_{i=1}^{m} p_{k,m}(j|\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T) + m p_{k,m}(j|\tilde{S}_1(m+1,k), \tilde{S}_2(m+1,k)^T) \]

\[ = mp_{k,m}(j) + m p_{k,m}(j) \quad \text{(4.37)} \]

The on-line updating formula is derived as follows:

\[ C_{j,m+1}(k) = \sum_{i=1}^{m+1} p_{k,m+1}(j|\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T)[\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T]^T S \sum_{i=1}^{m+1} p_{k,m+1}(j|\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T) \]

\[ \cong \sum_{i=1}^{m} p_{k,m}(j|\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T)[\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T]^T S \sum_{i=1}^{m+1} p_{k,m+1}(j|\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T) \]

\[ C_{j,m}(k)mp_{k,m}(j) + \sum_{i=1}^{m+1} p_{k,m}(j|\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T)[\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T]^T S \sum_{i=1}^{m} p_{k,m}(j|\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T) \]

\[ = mp_{k,m}(j) \quad \text{(4.38)} \]

\[ p_{k,m+1}(j) = \omega_{j,m+1}(k) = \frac{1}{m+1} \sum_{i=1}^{m+1} p_{k,m+1}(j|\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T) \]

\[ \cong \frac{1}{m+1} (p_{k,m}(j|\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T) + m p_{k,m}(j)) \]

\[ p_{k}^{\text{new}}(j|\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T) = \frac{p_{C}(\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T) C_{j,m+1}(k) \omega_{j,m+1}(k)}{\sum_{j=1}^{Q} p_{C}(\tilde{S}_1(i,k), \tilde{S}_2(i,k)^T) C_{j,m+1}(k) \omega_{j,m+1}(k)} \quad i = 1, 2, \ldots, m+1 \quad \text{(4.39)} \]

There are advanced on-line EM algorithms whose convergence has been proved, such as an algorithm proposed in [49]. Since the algorithm proposed in (4.38) is very simple and works fine, it will be used in this project. Anyway, there exist fast EM algorithms whose computation load is low.

**Remarks on the use of GMM**  In general, in speech processing, people usually models the spectrum or logarithm of the spectrum of power (or more often the Cepstrum) directly by GMM, without another assumption on the signal itself. The reasoning here, on the contrary, starts from a modeling on the signal and thus differs from the traditional speech processing. The modeling of signal will give rise to a possibly easy Bayesian inference in the later separation phase.
4.5.4 Separation phase: Weighted adaptive Wiener filtering

When \(s_1(n)\) and \(s_2(n)\) are stationary and (approximately) circular processes (i.e. with a Toeplitz covariance matrix), the basis which makes both covariance matrices diagonal is the discrete Fourier basis. Further, if the two sources are independent Gaussian processes, the best linear estimator of \(s_1(n)\) and \(s_2(n)\) from the mixture \(s(n)\) is obtained by Wiener filtering [33]:

\[
\hat{S}_1(m,k) = \frac{\sigma^2_1(k)}{\sigma^2_1(k) + \sigma^2_2(k)} S(m,k)
\]

\[
\hat{S}_2(m,k) = \frac{\sigma^2_2(k)}{\sigma^2_1(k) + \sigma^2_2(k)} S(m,k)
\]

(4.40)

where \(\sigma^2_i(k)\) forms the PSD shape of the signal \(s_i(n)\).

The Wiener filtering doesn’t apply in our case. It has to be extended in two ways. Firstly, the two signals in the Wiener filtering are independent. Thus it has to be generalized to correlated case. This can be done by Bayesian inference. Secondly, as the signals processed here are only locally stationary, the Wiener filtering with constant gain isn’t desirable anymore. An intuitive extension of the Wiener filtering is the adaptive Wiener filtering with time-varying gains, which are associated with the active components in GMM.

In a probabilistic formalism, the two sources can be estimated through a maximum likelihood (ML) estimate:

\[
(\hat{S}_1(m,k), \hat{S}_2(m,k)) = \arg \max_{\tilde{S}_1(m,k), \tilde{S}_2(m,k)} p(\tilde{S}(m,k)|\tilde{S}_1(m,k), \tilde{S}_2(m,k))
\]

(4.41)

The problem with ML approach is that there are multiple solutions since the system is under-determined. It is therefore nature to introduce the a posterior probability density function for the sources in a Bayesian formalism [33]:

\[
(\hat{S}_1(m,k), \hat{S}_2(m,k)) = \arg \max_{\tilde{S}_1(m,k), \tilde{S}_2(m,k)} p(\tilde{S}(m,k)|\tilde{S}_1(m,k), \tilde{S}_2(m,k))
\]

(4.42)

However, the MAP (maximum a posteriori) in (4.42) is not tractable. To get back the Wiener filtering case, a hidden random variable \(q\) (referred to as state variable) is introduced which is associated with the active Gaussian component in GMM, i.e., the Gaussian density from which the sources data are most likely generated.

\[
p(\tilde{S}(m,k)|\tilde{S}_1(m,k), \tilde{S}_2(m,k), q=j) = j p(q=j|\tilde{S}(m,k))
\]

(4.43)

\[
x \sum_{j=1}^{Q} p(\tilde{S}(m,k)|\tilde{S}_1(m,k), \tilde{S}_2(m,k), q=j) p(\tilde{S}_1(m,k), \tilde{S}_2(m,k)|q=j)
\]

Therefore the estimation of sources is carried out in two steps: estimate the current state by calculating \(p(q=j|\tilde{S}(m,k))\); constructing the filters by maximizing the posterior probability given the state \(j\).
State estimation \( p(q = j|\tilde{S}(m,k)) \) is the posterior probability of state variable given the observation \( \tilde{S}(m,k) \).

\[
\gamma_j(\tilde{S}(m,k)) = p(q = j|\tilde{S}(m,k)) \propto p(\tilde{S}(m,k)|q = j)p(q = j) \tag{4.44}
\]

When the active state \( j \) is given, \( \tilde{S}(m,k) \) is the sum of two correlated complex Gaussian variables with the joint distribution:

\[
\begin{bmatrix}
\tilde{S}_1(m,k) \\
\tilde{S}_2(m,k)
\end{bmatrix}
\sim CN(0,C_j(k)) \tag{4.45}
\]

where \( C_j(k) \) is given in (4.34). According to (4.24),

\[
\begin{bmatrix}
\text{Re}(\tilde{S}(m,k)) \\
\text{Im}(\tilde{S}(m,k))
\end{bmatrix}
\sim N(0,R_j(k)) \tag{4.46}
\]

\[
R_j(k) = \begin{bmatrix}
\sigma^2_{1,j}(k)/2 + \sigma^2_{2,j}(k)/2 + 2\sigma_{12,j} & 0 \\
0 & \sigma^2_{2,j}(k)/2 + \sigma^2_{1,j}(k)/2 + 2\sigma_{12,j}
\end{bmatrix}
\]

Thus,

\[
p(q = j|\tilde{S}(m,k)) \propto \omega_j(k)p_\omega([\text{Re}(\tilde{S}(m,k)),\text{Im}(\tilde{S}(m,k))])^T,R_j(k) \tag{4.47}
\]

Construction of the filters When the active state \( q \) is known, the problem (4.42) can be solved by extending the Wiener filtering.

\[ p(\tilde{S}(m,k)|\tilde{S}_1(m,k),\tilde{S}_2(m,k),q = j) \] is the probability of the observation \( \tilde{S}(m,k) \) when the two sources \( \tilde{S}_1(m,k),\tilde{S}_2(m,k) \) and the active Gaussian component are given. It can be derived as follows:

\[
p(\tilde{S}(m,k)|\tilde{S}_1(m,k),\tilde{S}_2(m,k),q = j) = \delta(\tilde{S}_1(m,k) + \tilde{S}_2(m,k) - \tilde{S}(m,k)) \tag{4.48}
\]

where

\[
\delta(x) = \begin{cases} 
1 & x = 0 \\
0 & \text{otherwise}
\end{cases} \tag{4.49}
\]

\[ p(\tilde{S}_1(m,k),\tilde{S}_2(m,k)|q = j) \] is the likelihood for the hidden \( q \) process and can be calculated straightforward:

\[ p(\tilde{S}_1(m,k),\tilde{S}_2(m,k)|q = j) = p_\omega([\tilde{S}_1(m,k),\tilde{S}_2(m,k)])^T,C_j(k)) \tag{4.50}
\]

where \( C_j(k) \) is given in (4.34).

Therefore, given the state \( j \), the posterior probability is:

\[
p(\tilde{S}_1(m,k),\tilde{S}_2(m,k)|\tilde{S}(m,k),q = j) \propto p(\tilde{S}(m,k)|\tilde{S}_1(m,k),\tilde{S}_2(m,k),q = j)p(\tilde{S}_1(m,k),\tilde{S}_2(m,k)|q = j) \tag{4.51}
\]
Taking the logarithm on both sides of (4.51) and substituting \( p(\tilde{S}(m,k)\mid \tilde{S}_1(m,k), \tilde{S}_2(m,k), q = j) \), \( p(\tilde{S}_1(m,k), \tilde{S}_2(m,k) \mid q = j) \) and \( p_0(\tilde{S}_1(m,k), \tilde{S}_2(m,k))^T, C_j(k) \) with (4.48), (4.50) and (4.35) respectively, we get:

\[
-\log(p(\tilde{S}_1(m,k), \tilde{S}_2(m,k)\mid \tilde{S}(m,k), q = j)) =
\left[ \tilde{S}_1(m,k), \tilde{S}_2(m,k) \right]^T C_j^{-1}(k) \left[ \tilde{S}_1(m,k), \tilde{S}_2(m,k) \right] - \log \delta(\tilde{S}_1(m,k) + \tilde{S}_2(m,k) - \tilde{S}(m,k)) + \text{Const.}
\] (4.52)

The MAP requires:

\[
\min_{\tilde{S}_1(m,k), \tilde{S}_2(m,k)} -\log(p(\tilde{S}_1(m,k), \tilde{S}_2(m,k)\mid \tilde{S}(m,k), q = j))
\text{s.t. } \tilde{S}_1(m,k) + \tilde{S}_2(m,k) = \tilde{S}(m,k)
\] (4.53)

Taking the derivative on both sides of (4.52) under the constraint in (4.53), the conditioned MAP estimator can be found. Since \( \tilde{S}_1(m,k), \tilde{S}_2(m,k), \tilde{S}(m,k) \) are all Gaussian, the conditioned MAP estimator is equivalent to the conditioned PM (Posterior Mean) estimator shown below:

\[
\tilde{S}_1(m,k) = E[\tilde{S}_1(m,k) \mid \tilde{S}(m,k), q = j] = \frac{\sigma_{1,j}^2(k) + \text{Re}(\sigma_{12,j}(k))}{\sigma_{1,j}^2(k) + 2 \text{Re}(\sigma_{12,j}(k)) + \sigma_{2,j}^2(k)} \tilde{S}(m,k)
\]

\[
\tilde{S}_2(m,k) = E[\tilde{S}_2(m,k) \mid \tilde{S}(m,k), q = j] = \frac{\sigma_{2,j}^2(k) + \text{Re}(\sigma_{12,j}(k))}{\sigma_{1,j}^2(k) + 2 \text{Re}(\sigma_{12,j}(k)) + \sigma_{2,j}^2(k)} \tilde{S}(m,k)
\] (4.54)

In (4.54), the MAP/PM estimator is conditioned on a known active state \( q \). When taking the active state \( q \) into account, the MAP and PM estimator of \( \tilde{S}_1(m,k), \tilde{S}_2(m,k) \) are different.

**MAP and PM estimator of two sources**

It is seen from (4.43), which gives the MAP estimator, that \( p(\tilde{S}_1(m,k), \tilde{S}_2(m,k)\mid \tilde{S}(m,k)) \) is maximized by picking up the Gaussian components with highest probability calculated in (4.47) as the active components. The estimation procedure then consists of the following two steps: find the active component \( j \) and estimate the two sources using (4.54).

A better estimator is PM since it avoids the hard decision of active components by assigning every component with a probability. The PM estimator is derived as follows:

\[
E[\tilde{S}_1(m,k) \mid \tilde{S}(m,k)] = \int_{\tilde{S}_1(m,k)} \tilde{S}_1(m,k) \int_{\tilde{S}_2(m,k)} p(\tilde{S}_1(m,k), \tilde{S}_2(m,k)\mid \tilde{S}(m,k)) d\tilde{S}_2(m,k) d\tilde{S}_1(m,k)
\]

\[
= \int_{\tilde{S}_1(m,k)} \tilde{S}_1(m,k) \int_{\tilde{S}_2(m,k)} \left( \sum_{j=1}^{q} \gamma_j(\tilde{S}(m,k)) \int_{\tilde{S}_1(m,k)} \int_{\tilde{S}_2(m,k)} (p(\tilde{S}_1(m,k), \tilde{S}_2(m,k)\mid \tilde{S}(m,k), q = j)) d\tilde{S}_2(m,k) d\tilde{S}_1(m,k) \right) d\tilde{S}_1(m,k)
\]

\[
= \sum_{j=1}^{q} \gamma_j(\tilde{S}(m,k)) \int_{\tilde{S}_1(m,k)} \int_{\tilde{S}_2(m,k)} (p(\tilde{S}_1(m,k), \tilde{S}_2(m,k)\mid \tilde{S}(m,k), q = j)) d\tilde{S}_2(m,k) d\tilde{S}_1(m,k)
\]

\[
= \sum_{j=1}^{q} \gamma_j(\tilde{S}(m,k)) E[\tilde{S}_1(m,k) \mid \tilde{S}(m,k), q = j]
\] (4.55)
Thus,

$$E[\tilde{S}_i(m,k)\big|\tilde{S}(m,k)] = \sum_{j=1}^{Q} \gamma_j(\tilde{S}(m,k))E[\tilde{S}_i(m,k)\big|\tilde{S}(m,k),q=j] \quad i=1,2 \quad (4.56)$$

where $\gamma_j(\tilde{S}(m,k))$ is computed in (4.44), $E[\tilde{S}_i(m,k)\big|\tilde{S}(m,k),q=j] E[\tilde{S}_2(m,k)\big|\tilde{S}(m,k),q=j]$ are derived in (4.54).

According to (4.56), the first step of PM estimation is to calculate the posterior probability $\gamma_j(\tilde{S}(m,k))$. The second step is to construct the weighted adaptive Wiener filter with weights $\gamma_j(\tilde{S}(m,k))$, which depend on the observation, and filter the observation with constructed adaptive filter.

The diagram of the novel algorithm is thus depicted as below, where $S_i(1), S_i(2), \ldots, S_i(L) i=1,2$ are the coefficients of basis functions, $\theta_i(1), \theta_i(2), \ldots, \theta_i(L)$ is the parameters of PDF model. Note that this is a very general diagram of the algorithm: DFT is only one instance of basis functions; GMM is one kind of PDF model and weighted adaptive Wiener filtering is a special result from Bayesian inference.

![Figure 4-13 General diagram of proposed source separation algorithm](image)

### 4.5.5 Performance of proposed source separation algorithm

The proposed algorithm is simulated to evaluate the performance. The test setup is shown in the following table:
4.5 Weighted adaptive Wiener filter: A complicated scheme

Table 4-5 Simulation setup for source separation algorithm

<table>
<thead>
<tr>
<th>The windowing</th>
<th>Hanning window</th>
</tr>
</thead>
<tbody>
<tr>
<td>The length of frame/DFT</td>
<td>512 samples</td>
</tr>
<tr>
<td>The length of over-lapping</td>
<td>half the length of frame</td>
</tr>
<tr>
<td>Number of components in GMM</td>
<td>3</td>
</tr>
<tr>
<td>Number of training frames</td>
<td>1000</td>
</tr>
<tr>
<td>Number of separation frames</td>
<td>100</td>
</tr>
</tbody>
</table>

The measure of source separation performance is defined as normalized test error:

\[ \xi_i = \frac{\|\hat{s}_i(n) - s_i(n)\|}{\|s_i(n)\|} \quad i = 1, 2 \]  

(4.57)

**Simulation 1** In this simulation, the two sources \( s_1(n) \) and \( s_2(n) \) are two zero-mean correlated white Gaussian noises. Different cases are studied in order to find out how the proposed algorithm performs in various situations.

When the two sources \( s_1(n) \) and \( s_2(n) \) are two zero-mean correlated white Gaussian noises with the covariance matrix \([2, 1.2; 1.2, 2]\), which corresponds to 60% correlation, the results are very satisfactory as shown below:

\[ \xi_1 = 0.2053 \]

Figure 4-14 Results of proposed source separation algorithm when the sources are two correlated Gaussians (1)

When the two correlated Gaussian sources have very different variance with the covariance matrix \([2, 2.4; 2.4, 8]\), still corresponding to 60% correlation, the separation performance is different from above as shown below:

\[ \xi_2 = 0.2030 \]
Chapter 4. Source Separation for Feedback Cancellation

\[ \delta_1 = 0.3483 \]

Figure 4-15 Results of proposed source separation algorithm when the sources are two correlated Gaussians (2)

It is seen that the estimation of \( s_1(n) \) is worse than the estimation of \( s_2(n) \). It means for two sources with different energy, the dominant one will be estimated better. This can be explained easily. Due to the constraint used in (4.53) in the separation phase,

\[ s(n) = s_1(n) + s_2(n) = \hat{s}_1(n) + \hat{s}_2(n) \]  

(4.58)

Therefore,

\[ s_1(n) - \hat{s}_1(n) = \hat{s}_2(n) - s_2(n) \]

\[ \| s_1(n) - \hat{s}_1(n) \| = \| s_2(n) - \hat{s}_2(n) \| \]  

(4.59)

\[ \| s_1(n) - \hat{s}_1(n) \| \gg \| s_2(n) - \hat{s}_2(n) \| , \text{i.e., } \delta_1 \gg \delta_2 \]

Thus the dominant source is always estimated better, the more dominant the better estimated. This is usually desired because the interesting signal is often stronger.

When the two sources have negative correlation, things are totally different. Suppose the two correlated Gaussian noises have the variance matrix of \([2,-1.2; -1.2,2]\), corresponding to a negative 60% correlation, the separation results are illustrated in the following figures:

\[ \delta_1 = 0.8030 \]

\[ \delta_2 = 0.7977 \]

Figure 4-16 Results of proposed source separation algorithm when the sources are two correlated Gaussians (3)
It is obvious that the performance degrades a lot in this case. The reason is that the proposed adaptive filtering is based on Wiener filtering which is phase blind since it is only based on power spectral densities without taking phase into account. An intuitive way to explain this is that: when $s_1(n)$ is positive (above the zero mean), $s_2(n)$ tends to be negative (below the zero mean). Therefore, in the sum of the two sources, some amount is actually cancelled. The amount of cancellation depends on the correlation between the two sources and the variances of the two sources. An extreme case is that $s_1(n)$ and $s_2(n)$ have the covariance matrix $[2,-2;-2,2]$, i.e., $s_1(n)$ is totally dependent on $s_2(n)$ and $s_1(n) + s_2(n) = 0$. In this case the observation is always 0, separation is impossible.

To estimate how much is cancelled in the sum, additional information has to be obtained during training, such as phase information. Suppose besides the probability density model in each frequency bin, there is an additional phase model, modeling the phase difference between the two signals. (The phase difference is easier to model since it represents the phase shift in each frequency bin resulted from forward-path in hearing aids combined with feedback path and thus is more static or stable than the rapidly changing absolute phases of the two sources.). The negative correlation problem can be solved easily since two equations (frequency response and phase) with two unknowns are deterministic. Or a better way can be obtained by Bayesian inference such as:

\[
\begin{align*}
(\tilde{S}_1(m,k), \tilde{S}_2(m,k)) &= \arg \max_{\tilde{S}_1(m,k), \tilde{S}_2(m,k)} p(\tilde{S}(m,k)|\tilde{S}_1(m,k), \tilde{S}_2(m,k)) p(\angle \tilde{S}(m,k) | \angle \tilde{S}_1(m,k) - \angle \tilde{S}_2(m,k)) \\
&= (4.60) \\
\text{Or} \\
(\tilde{S}_1(m,k), \tilde{S}_2(m,k)) &= \arg \max_{\tilde{S}_1(m,k), \tilde{S}_2(m,k)} p(\tilde{S}_1(m,k), \tilde{S}_2(m,k)|\tilde{S}(m,k)) p(\angle \tilde{S}_1(m,k) - \angle \tilde{S}_2(m,k) | \angle \tilde{S}(m,k)) \\
&= (4.61)
\end{align*}
\]

Due to the limit of time, phase modeling is skipped in this project. However, it is vital to improve the performance of proposed source separation algorithm.

**Simulation 2** In this simulation, more complicated signals are tested.

In the first test, the two sources are generated from a Gaussian Mixture Model:

\[
p(s_1(n),s_2(n)) = \sum_{j=1}^{3} \omega_j p_\mathcal{G}([s_1(n), s_2(n)]^T, C_j)
\]

where

\[
\omega = [0.5,0.2,0.3] \\
C_1 = [2,1.7;1.7,4] \\
C_2 = [10,2.5;2.5,16] \\
C_3 = [4,1.8;1.8,1]
\]

The results are shown below:
The results are satisfactory though the signals are only locally stationary. It shows that the training phase does obtain enough prior knowledge on the structure of sources.

In the second test, correlated 10th order AR processes are used. $s_1(n)$ is a 10th order AR process generated by a special set of coefficients that causes severe violations of assumptions in the end of section 4.5.2. $s_2(n)$ is $s_1(n)$ filtered by impulse response of a measured feedback path and amplified by 500 so that $s_1(n)$ and $s_2(n)$ have comparable variances.
The results are very satisfactory though in section 4.5.2., it is mentioned that this set of coefficients violates the assumptions severely. Moreover, the negative correlation problem probably happens in some frequency bins due to the phase shift of feedback path. However, the overall performance is not compromised at all. This example shows that the proposed algorithm is very robust.

**Simulation 3** In this simulation, real speech is tested. $s_1(n)$ is a male speech. $s_2(n)$ is $s_1(n)$ filtered by impulse response of the measured feedback path shown in Figure 4-18 and amplified by 40 dB so that $s_1(n)$ and $s_2(n)$ have comparable variances. The results are shown below:
Chapter 4. Source Separation for Feedback Cancellation

\[ \xi_1 = 0.1463 \]

\[ \xi_2 = 0.1654 \]

speech spectrogram

estimated speech spectrogram
4.5 Weighted adaptive Wiener filter: A complicated scheme

The performance is even better than white noise case in simulation 1. This is because speech is very structured signal. As long as the structure is captured by GMM, the estimation is very accurate.

Compare the original spectrograms with the estimated spectrograms, it is seen that the original signal is cut off at 4 kHz. Though the estimated signal is not sharply cut off there due to the high sampling rate (11025 Hz) the frequency contents are small enough above 4 kHz.

A pattern is found in the estimated spectrograms: horizontal broken lines are located at some evenly spaced frequencies. This indicates the failure of modeling at those frequencies. The reason is that voiced speech shows strong tonal characteristics at harmonic frequencies. As we all know, the DFT coefficients of a sinusoid signal are either 0 or a constant value. Thus GMM model doesn’t work.

The number of Gaussian components Simulations are carried out using the same speech and the speech filtered by the feedback path as in simulation 3. The training is based on the first 1000 frames. The successive 500 frames are used as test data and normalized error is calculated for the test data. The results are shown below:
Table 4-6 Performance as a function of the number of Gaussian components

<table>
<thead>
<tr>
<th>Number of Gaussian components</th>
<th>Normalized error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\zeta_1 = 0.1433, \zeta_2 = 0.1628$</td>
</tr>
<tr>
<td>2</td>
<td>$\zeta_1 = 0.1226, \zeta_2 = 0.1438$</td>
</tr>
<tr>
<td>3</td>
<td>$\zeta_1 = 0.1223, \zeta_2 = 0.1400$</td>
</tr>
<tr>
<td>4</td>
<td>$\zeta_1 = 0.1290, \zeta_2 = 0.1465$</td>
</tr>
<tr>
<td>5</td>
<td>$\zeta_1 = 0.1306, \zeta_2 = 0.1483$</td>
</tr>
<tr>
<td>6</td>
<td>$\zeta_1 = 0.1324, \zeta_2 = 0.1503$</td>
</tr>
<tr>
<td>9</td>
<td>$\zeta_1 = 0.1325, \zeta_2 = 0.1504$</td>
</tr>
<tr>
<td>12</td>
<td>$\zeta_1 = 0.1340, \zeta_2 = 0.1522$</td>
</tr>
</tbody>
</table>

Figure 4-20 shows that the test error drops down and later increases as the number of Gaussian components increases. The drop indicates the inadequacy of Gaussian components in describing the power spectral densities. The rising up is a consequence of over-fitting. The optimal number for speech and the speech filtered by the impulse response of a measured feedback path is 3.

The length of frame/DFT When the number of GMM is fixed as 3, the length of frame is investigated to find the relation between the separation performance and the spectrum resolution. Simulations are carried out using the same speech and the speech filtered by the feedback path as in simulation 3. The training is based on the first 1000 frames. The successive 500 frames are used as test data and normalized error is calculated for the test data. The results are shown below:
Table 4-7 Performance as a function of frame length

<table>
<thead>
<tr>
<th>Length of frame</th>
<th>Normalized error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1024 samples</td>
<td>$\varsigma_1 = 0.1032, \varsigma_2 = 0.1203$</td>
</tr>
<tr>
<td>512 samples</td>
<td>$\varsigma_1 = 0.1223, \varsigma_2 = 0.1400$</td>
</tr>
<tr>
<td>256 samples</td>
<td>$\varsigma_1 = 0.1495, \varsigma_2 = 0.1468$</td>
</tr>
<tr>
<td>128 samples</td>
<td>$\varsigma_1 = 0.1820, \varsigma_2 = 0.1963$</td>
</tr>
<tr>
<td>64 samples</td>
<td>$\varsigma_1 = 0.1482, \varsigma_2 = 0.2890$</td>
</tr>
</tbody>
</table>

Figure 4-21 Normalized test error as a function of the frame length

Generally speaking, the performance is better with a finer spectrum, i.e., longer frames. Since time domain modeling can be regarded as a special frequency-domain modeling when the frame length equals to 1, the conclusion infers that time-domain modeling is worse than frequency-domain modeling.

4.5.6 Proposed algorithm in DCT domain

As shown in Figure 4-13, DFT is only one kind of basis functions in proposed separation algorithm. Another widely used basis function is DCT (discrete cosine transform). It is tried also in this project because it has advantages over DFT. DCT is defined as follows:

$$X(k) = w(k) \sum_{n=1}^{L} x(n) \cos\left(\frac{\pi(2n-1)(k-1)}{2L}\right), \quad k = 1, 2, \ldots, L$$  \hspace{1cm} (4.62)

where
Firstly, the coefficients of DCT are real which simplifies the algorithm proposed above and thus reduces the computation load. Secondly, DTC provides a set of basis vectors that, together, are a good approximation to the KLT whose coefficients are uncorrelated. Indeed, for a stationary zero-mean, first-order Markov process that is deemed to be sufficiently general in signal-processing studies, the DCT is asymptotically equivalent to the KLT, both as the sequence length increases and also as the adjacent correlation coefficient tends toward unity. [10]

The assumptions (4.27) and (4.28) are also desired in DCT domain to validate the derivation in section 4.5.3 and 4.5.4. Fortunately, through numerous computer simulations, these assumptions still hold approximately and even better than in DFT domain.

The source separation performance in the DCT domain is found to be as good as in DFT domain. The reason is that for most signals the assumptions in (4.27) and (4.28) in DFT domain hold as well as in DCT domain. However, DCT significantly reduces the computation load and is therefore preferred. In hearing aids, the computation advantage is compromised since FFT is already an existing module shared by many other algorithms but DCT is not yet.

### 4.5.7 Incorporate source separation into feedback cancellation

As seen in section 4.5.5, the proposed source separation algorithm works better when the feedback signal is strong. This is a desired feature since it is exactly when the feedback canceller needs help.

Assume that the characteristics of the input signal don’t change abruptly, for example, from pure speech to music. When the modeled density function in the spectrum is accurate, the estimation of the input signal should be helpful to the feedback cancellation.

#### Strategies of incorporating

Two strategies are proposed to incorporate source separation information into the feedback canceller. Both are trying to constrain the estimation of feedback signal from the feedback canceller by pulling it towards the estimation from source separation. Define the information provided to the feedback canceller from source separation \(ss(n)\) (see Figure 4-9) as the estimation of \(e(n)\), i.e.,

\[
ss(n) = \hat{e}(n)
\]  

(4.63)

**Strategy I**  The first strategy is to constrain the adaptation of the adaptive filter. It is similar to the constrained feedback canceller in section 2.4 and 2.5. For block version constrained LMS, the cost function is:

\[
J(n) = \sum_{n=0}^{L-1} |e(n)|^2 + \beta \sum_{k=0}^{p-1} [\hat{w}_k(n) - \hat{w}_k(0)]^2 + \alpha \sum_{n=0}^{L-1} (e(n) - ss(n))^2
\]

(4.64)

where \(L\) is the number of samples per block, \(p\) is the order of adaptive filter. \(\hat{w}_k(0)\) is the coefficient of the reference filter. \(\alpha, \beta\) are step sizes. When \(\beta = 0\), the LMS is normal, not constrained by the reference filter. The minimization of (4.64) leads to a new updating formula:
\[ \hat{w}_k(n+1) = \hat{w}_k(n) - \mu \beta [\hat{w}_k(n) - \hat{w}_k(0)] + \mu \sum_{m=0}^{L-1} d_{m-k}(n)[(1 + \alpha) e_m(n) - \alpha ss_m(n)] \] (4.65)

where \( d_{m-k}(n) \) is the input to the LMS filter.

**Strategy II**  
The second strategy constrains the output of feedback canceller. It is equivalent to constraining \( e(n) \):

\[ e(n) = s(n) - (\lambda v(n) + (1 - \lambda)(s(n) - ss(n))) \] (4.66)

\( \lambda \) satisfies \( 0 \leq \lambda \leq 1 \), determining how strongly the estimation of feedback canceller is constrained by the source separation.

### 4.5.8 Simulations results

Simulations are made to evaluate the performance of the two strategies. The evaluation is based on a compare between the finely tuned feedback canceller and the same feedback canceller assisted by the proposed source separation algorithm using strategy I and II respectively. The setup of simulation is show in the following table:

| Table 4-8 Simulation setup for proposed strategies

<table>
<thead>
<tr>
<th>Test signals:</th>
</tr>
</thead>
<tbody>
<tr>
<td>flute (30 seconds, 16 kHz sampling rate)</td>
</tr>
<tr>
<td>male speech (30 seconds, 16 kHz sampling rate)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feedback canceller:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A standard feedback canceller with filtered-X modified LMS adaptation algorithm tuned to best performance</td>
</tr>
<tr>
<td>Advanced feedback canceller developed by GN Resound tuned to best performance</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Block size of feedback canceller and source separation:</th>
</tr>
</thead>
<tbody>
<tr>
<td>128 samples</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feedback path:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measured in real life including sudden change resulted from environmental change.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Forward path of hearing aids:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A delay (8 samples, i.e., 0.5 ms) + Gain (20 dB) for standard feedback canceller</td>
</tr>
<tr>
<td>A delay (8 samples, i.e., 0.5 ms) + Gain (29 dB) for GN Resound’s feedback canceller</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Performance measure of feedback canceller:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additional stable gain</td>
</tr>
</tbody>
</table>
The two type of input signal are very representative. Speech is the main class of signals hearing aids process. The flute is very tonal, representing the case when feedback canceller encounters bias problem severely. The diagram of feedback canceller from GN Resound is shown in Figure 2-6. The standard feedback canceller is similar to GN Resound’s feedback canceller except that frozen IIR filter and constraints on adaptive filter are removed (see Figure 2-6). The block size is chosen as 128 samples which is a compromise between the fine structure of spectrum required by source separation as pointed in section 4.5.5 and realistic parameters in hearing aids (24 or 56 samples). The large block size required by source separation doesn’t prohibit the use of source separation from practical use since this problem can be solved by over-lapping techniques in spectrum estimation. For simplicity this is not implemented in the simulation. The performance measure is ASG (see section 2.6) because filtered-X is used in both cancellers emphasizing the estimation on critical bands where feedback whistling usually occurs. The forward path is a simply delay plus gain. The gain is selected above the stable gain of hearing aids without feedback cancellation and approaches the upper bound performance of the feedback canceller. With this gain, it is able to show how the whistling occurs and feedback canceller removes it gradually. GN Resound has supplied a set of feedback paths measured every 60ms in real life. These feedback paths are interpolated to provide a smooth time-varying feedback path for the simulation. It is so realistic that it includes the feedback path when putting telephone close to the ear and so on. The feedback path used in the simulation is shown in Figure 4-22.

In the first 1000 frames, a batch version EM algorithm (see Eq. (4.36)) is used to estimate the initial parameters in GMM. Thus there will be no assistance to the feedback canceller. After that, the on-line EM algorithm is performed frame by frame (see Eq.(4.38)) to tolerate the change of statistics of feedback signal due to the change of feedback path, and provide the output to the feedback canceller. In the following simulations, the results will show ASG only after the frame 1000.
4.5 Weighted adaptive Wiener filter: A complicated scheme

![Frequency response of feedback path over time](image)

**Figure 4-22 Frequency response of feedback path over time**

**GN Resound’s feedback canceller**  The parameters are listed in the following table when the performance is tuned to the best. All the parameters and module names are referred to in Eq. (2.20) and Figure 2-6.

<table>
<thead>
<tr>
<th>Table 4-9 Optimal parameters for GN Resound’s feedback canceller</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adaptive FIR filter:</strong></td>
</tr>
<tr>
<td>12 orders, 128-sample block size</td>
</tr>
<tr>
<td>( \mu = 2.25e - 5, \beta = 0.002, \lambda = 0.995, c = 1e - 5 )</td>
</tr>
<tr>
<td><strong>Frozen IIR filter:</strong></td>
</tr>
<tr>
<td>6 orders</td>
</tr>
<tr>
<td><strong>High pass filter (filtered-X):</strong></td>
</tr>
<tr>
<td>([0.6667 - 0.3333])</td>
</tr>
<tr>
<td><strong>DC filter:</strong></td>
</tr>
<tr>
<td>([1 -1])</td>
</tr>
</tbody>
</table>

Delay in the adaptive path, IIR filter coefficients and reference FIR filter are all found optimally during initialization. The optimal step size of LMS \( \mu = 2.25e - 5 \) is found by MMSE (minimize mean square error) criterion through numerous simulations.

The following figure, showing ASG with different step size \( \mu \), serves as an example to illustrate the behavior of feedback canceller with different step sizes. It is seen that bigger step size leads to fast convergence when the feedback path changes (compare with Figure 4-22). However, when the step
size is too big, the feedback canceller diverges (when $\mu = 5e-5$ which is not shown in the following figure, it diverges).

![Figure 4-23 Example showing how the feedback canceller behaves with different step sizes](image)

**Strategy I**  
The scheme is described in (4.65). For GN Resound’s feedback canceller, it is rewritten as:

$$\hat{w}_k(n + 1) = \hat{w}_k(n) - \beta [\hat{w}_k(n) - \hat{w}_k(0)] + \mu \sum_{m=0}^{k-1} v_{m-k}(n) [(1 + \alpha) e_m^f(n) - ass^f(n)]$$  

(4.67)

where the superscript $f^\prime$ denotes the signal filtered by the high pass filter (filtered-X).

For flute signal, the best $\alpha$ is found to be 5. The result is shown below:

![Figure 4-24 Performance evaluation of strategy I in GN Resound’s feedback canceller (flute input)](image)
The big fluctuations of the two lines are mainly resulted from sudden changes in the feedback path (compare with Figure 4-22). The quick falling down shows that the changes are so abrupt that feedback canceller can’t catch up. The rise up shows that feedback canceller adapts according to the current feedback path and prevents the system from instability.

When comparing the two curves, it is seen that the assist of source separation can alleviate the quick drop of ASG by 1~2 dB (around 2 dB at the dip between frame 1500 to 2000, 1 dB at the dip between frame 2000 to 2500). This is very valuable since it shows source separation does help when the system is approaching to instability, which is the most dangerous case.

Sometimes the assistance even helps the feedback canceller to converge faster and better (around 6 dB more ASG at the peak between frame 2000 to 2500). However, considering the overall performance, the assistance doesn’t help much with increasing the ASG when the feedback canceller recovers from misbehavior.

The reason why source separation can alleviate the dip in Figure 4-24 can be explained by investigating the estimation of feedback signal $f(n)$.

![Time domain analysis](image)

**Figure 4-25** Compare between real feedback signal and estimations from feedback canceller and source separation

Figure 4-25 shows the feedback signal estimations in the time segment corresponding to the location of ASG dip of feedback canceller with the assist of source separation. It shows clearly why source separation assisted feedback canceller has a lower ASG dip between frame 2000 and 2500. A sudden change of the real feedback signal (red) is observed from sample 2.88e5 to 2.89e5 corresponding to an abrupt change in the feedback path. Since the feedback signal is getting larger and larger, the feedback path change is probably caused by picking up the phones or putting hands close to the ear, both are the usual cases when whistle occurs.

In the portion from sample 2.89e5 to 2.90e5, the estimation of feedback canceller (blue) deviates farther and farther away immediately after the change. However, source separation follows the change in the feedback signal closely and gives much better estimation, which helps the feedback
canceller return to the right track by exploiting (4.65). This also shows the effectiveness of our own on-line EM algorithm, derived in section 4.5.3, works since it can track the change of statistics of feedback signal resulted from the change of feedback path. In the portion from sample 2.90e5 to 2.91e5, it is clearly seen that the estimation of feedback canceller is pulled towards the real feedback signal. However, after a period, the estimation from source separation doesn’t help anymore and even follows the feedback canceller instead of pulling it further.

This illustrates the mutual relation between feedback canceller and source separation. On the one hand, source separation is based on probability density function, which doesn’t change as fast as feedback canceller and won’t deviate from right track so easily. Therefore, it can help feedback canceller in the beginning. However, on the other hand, since all the information source separation module gets is from feedback canceller, they will finally give similar estimations of feedback signal.

A new idea is inspired by the above analysis: Stop or slow down the adaptation of the parameters in source separation (e.g., GMM) when feedback canceller by other means, such as whistle detection, is judged to be misbehaving. In this way, source separation will not be misled by feedback canceller anymore and the help from source separation can be lengthened. At the same time, the step size $\alpha$ could be proportional to the absolute difference between estimation from the feedback canceller and source separation to strengthen the helping force. Due to the limit of time in this project, this new idea is not implemented.

It is also noted that the overall ASG curve with source separation is smoother than without. This is an advantage in the sense of sound quality. The reason is still that source separation is based on probability estimation which isn’t sensitive to sudden changes.

For male speech input, the best $\alpha$ is found to be 2.3. The result is shown below:

![Figure 4-26 Performance evaluation of strategy I in GN Resound’s feedback canceller (speech input)](image)

Compared with Figure 4-24, the benefit is not as large. Especially when the feedback canceller goes back to the right track, the assistance actually hinders the recovery. Similar to Figure 4-24, the
assistance of source separation also hinders the ASG of feedback canceller from dropping down too much. The two observations are both because of the insensitivity of the probability density function to sudden changes. This “lag” effect explains why the assisted feedback canceller usually has smaller ASG in the peaks and higher ASG in the dips.

Another new idea to address the lower peaks problem is to stop or weaken the influence of source separation on feedback canceller when it is recovering from misbehavior, so that it is not hindered by the lag source separation. A direct way of implementing this idea is to decrease the step size $\alpha$ in (4.65) when the feedback canceller is going back to the right track. Due to the limit of time in this project, this new idea is not implemented.

The flute input represents the case when feedback canceller encounters bias problem severely. Therefore, it leaves more room to improve. Feedback canceller normally works fine with speech. That is possibly why the benefit for speech input is not as much as for flute signals.

**Strategy II** The second strategy is to control the estimation of feedback canceller instead of constraining the coefficients of adaptive filter directly.

For both flute signal and male speech, the performance is not exciting, and even a little worse than the original feedback canceller as illustrated below:

![Figure 4-27 Performance evaluation of strategy II in GN Resound’s feedback canceller (flute input)](image-url)
Chapter 4. Source Separation for Feedback Cancellation

The reason is probably that in strategy II the control of estimation of feedback signal has less influence on the adaptive filter than in strategy I since it is in an indirect way. When feedback canceller misbehaves, the source separation receives wrong information and misbehaves before it is able to help feedback canceller. Then the constraint of feedback signal estimation makes no sense and even degrades the performance.

**Standard feedback canceller** The parameters are listed in the following table when the performance is tuned to the best. All the parameters and module names have the same meaning as used in Eq. (2.20) and Figure 2-6.

<table>
<thead>
<tr>
<th>Table 4-10 Optimal parameters for standard feedback canceller</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adaptive FIR filter:</strong></td>
</tr>
<tr>
<td>16 orders, 128-sample block size</td>
</tr>
<tr>
<td>$\mu = 1e^{-6}, \lambda = 0.995, c = 1e^{-5}$</td>
</tr>
<tr>
<td><strong>High pass filter (filtered-X):</strong> [0.6667 -0.3333]</td>
</tr>
<tr>
<td><strong>DC filter:</strong> [1 -1]</td>
</tr>
</tbody>
</table>

*Optimal delay in the adaptive path is found by an algorithm similar to initialization for GN Resound’s feedback canceller to give a best match between feedback path and adaptive path. The optimal step size of LMS $\mu = 1e^{-6}$ is found by MMSE (minimize mean square error) criterion through numerous simulations.*
4.5 Weighted adaptive Wiener filter: A complicated scheme

Strategy I  The scheme is described in (4.65). For standard feedback canceller, it is rewritten as:

\[ \hat{w}_k(n+1) = \hat{w}_k(n) + \mu \sum_{m=0}^{k-1} v_{m-k}^f(n)[(1 + \alpha)e_m^f(n) - \alpha s_s^f(n)] \]  

(4.68)

where the superscript \( f \) denotes the signal filtered by the high pass filter (filtered-X).

For flute signal, the best \( \alpha \) is found to be 15. The result is shown below:

![Graph showing performance evaluation of strategy I in standard feedback canceller (flute input)](image)

Figure 4-29 Performance evaluation of strategy I in standard feedback canceller (flute input)

Similar to the findings in GN Resound’s feedback canceller, the source separation assisted feedback canceller has elevated dips and floors in the ASG curves. It also has suppressed peaks. When \( \alpha \) is larger, the elevation of dips and floors, and suppression of peaks are heavier, which proves the effectiveness of the new idea proposed just now. These properties are both attributed to the “lag” effect of source separation too.

For male speech input, the best \( \alpha \) is 5. The result is shown below for different \( \alpha \) :
Additional stable gain

**Figure 4-30 Performance evaluation of strategy I in standard feedback canceller (speech input)**

From Figure 4-30 the similar conclusions can be drawn as for the flute input. The reason is similar too.

**Strategy II** Similar to the findings in GN Resound’s feedback canceller, for both flute signal and male speech, the performance is not exciting as illustrated in the figures below. The reason is also similar. This shows strategy II is not an effective way of incorporating source separation in feedback canceller.

**Figure 4-31 Performance evaluation of strategy II in standard feedback canceller (flute input)**
Conclusions of simulations Two strategies were proposed when incorporating the source separation algorithm in the feedback canceller. The first strategy is successful in preventing the system from instability. It shows two advantages: elevating the dips in ASG curve, which may enhance the stability of the feedback system, and smoothing the curve, which might improve the sound quality. These two merits are both attributed to the “lag” effect of source separation which is based on probability density function.

The mutual influence of feedback canceller and source separation is the key to utilize the proposed algorithm in feedback cancellation system. Two new ideas are proposed, which are not implemented due to the limit of time in this project.

4.6 Summary

In this chapter, source separation techniques for feedback cancellation are investigated. The aim of using the source separation techniques in feedback cancellation is to improve the performance of existing feedback canceller. Instead of replacing it, source separation is served as an additional module in the hearing-aid systems.

The usage of source separation techniques can be divided into two types: indirect way and direct way. The latter separates the mixture signal explicitly, while the former only takes advantage of the ideas originating from source separation.

As an indirect way of using source separation techniques in feedback cancellation system, an ICA-based LMS is investigated. It only gives marginal benefit due to the assumption of independence is violated in feedback problem.
Direct usage of source separation techniques includes blind and semi-blind source separation techniques. Feedback cancellation with linear prediction is an example of using blind source separation techniques. Though the benefit is not found, it is still an interesting idea under investigation by some researchers today.

A novel algorithm, weighted adaptive Wiener filter for under-determined source separation problem with correlated sources, is proposed, which is in fact a semi-blind source separation technique. The proposed algorithm is very successful in separating two correlated sources with only one microphone especially when the two signals have comparable variance.

Two strategies are proposed to incorporate the new algorithm in feedback cancellation. They are implemented on two feedback canceller, one of which is the state of art in feedback cancellation system, to evaluate the performance. The first strategy is proved to be successful in enhancing the stability of the feedback system. The common worry about computation load is relieved by utilizing training phase only in critical frequency bands, using 3 components proved to be optimal in GMM for speech signals, and adopting a simple on-line EM algorithm developed in this thesis.
Chapter 5 Conclusion

This chapter summarizes the work done in this project and gives the thoughts on future research in source separation for feedback cancellation in hearing aids.

5.1 Work done in this thesis

The thesis is the first attempt to use source separation techniques for feedback cancellation in hearing aids. The work that has been done during the thesis is shown in the following aspects:

- The feedback suppression techniques are reviewed thoroughly. The fundamental limitation of feedback cancellation, bias problem, is clarified by both theoretical deduction and simulations. The state of art in feedback cancellation is also elaborated.

- The source separation techniques including blind source separation and semi-blind source separation are reviewed. This work has been difficult due to the large number of available techniques. The various techniques scattered in papers or books are categorized and briefly introduced. The emphasis is placed on under-determined case and correlated sources case because of their relevance with this project. Through this work, the main direction of using source separation techniques in feedback cancellation is determined, i.e., blind source separation algorithms with general but strong assumptions or semi-blind source separation algorithms with training phase to obtain enough prior knowledge.

- The Gaussian statistical model for DFT coefficients (complex random values) of a random process is built up. This assumption is used very widely, but very little literature has given such detailed description. Strict mathematical conditions for the assumption to hold are given. The statistical relation between the DFT coefficients of two random processes is also investigated. All the assumptions are verified by numerous computer simulations. It was found that the assumptions hold in most cases. But exceptions do exist.

- The possible usage of source separation in feedback cancellation is studied. An indirect usage base on ICA is tried. It provides marginal benefit. A direct way using linear prediction is also tried out. Though the performance is not as good as expected, the method is still worth investigating as discussed in [9].

- A novel algorithm for source separation with under-determined case and correlated sources is proposed. The simulation shows that the algorithm works well especially when the two sources have comparable energy. The problems with the algorithm still exist. The fundamental reasons are the loss of phase information and the constraint on the sum of the two signals.
Two strategies, which are the first attempt to use source separation to assist the feedback canceller, are proposed to embody the source separation techniques in feedback canceller. The first one is proved to be effective in enhancing the system stability. Since the reference systems are the state of art in feedback cancellation and a standard feedback canceller after being carefully tuned, the benefit is valuable and generally applicable. Two new ideas to improve the performance are also proposed in this thesis. Moreover, the computation load is paid attention to. It can be reduced by utilizing training phase only in critical frequency bands, using 3 components proved to be optimal in GMM for speech signals, and adopting a simple on-line EM algorithm developed in this thesis.

5.2 Future research

As a result of the work presented in this thesis, the following future work is suggested:

The temporal structure of the source, such as correlation of the samples, is not utilized enough in the proposed source separation algorithm. It will be beneficial to take this into account for both the source separation algorithm itself and the feedback canceller assisted by source separation.

To address the problem with negative correlation in weighted adaptive Wiener filtering algorithm, acoustic phase modeling has to be included. Phase information is vital for the separation performance and also helps to improve the sound quality. Due to the limit of time, this part is skipped. But several thoughts have come up. The first is to use a phase vocoder to model the phase change over time and derive a new separation algorithm based on Bayesian framework. The second thought is to use Hilbert transform to find the phase relation between the two sources. The phase model should model the phase difference between the two sources in each frequency bin instead of the absolute phases for the two sources since the phase difference represents the phase shift at each frequency bin resulted from forward-path in hearing aids combined with feedback path and is thus more static or stable than the rapidly changing absolute phases of the sources.

The modeling of probability density function in weighted adaptive Wiener filtering is GMM. More advanced models such as LMM (Laplacian Mixture Model) or HMM (Hidden Markov Model) should give better performance in obtaining the accurate structure of the signal. For example, in HMM, a state dependent mixture of Gaussians components should be superior to a single state model GMM.

The GMM model has a limitation. When in the context of audio processing, we may observe the same sound corresponding to a similar PSD shape, repeated at different amplitudes and time indexes. There has to be as many Gaussian components as there are different possible amplitudes to model the PSD shapes, although they correspond to the same sound. This is quite restrictive. A more elaborate model: the Gaussian scaled mixture model (GSMM), can be used in order to separate the variance shape (PSD), and the amplitude information (gain factor).[33] Since the gain vector is estimated over all the frequency bins, that is to say, all the frequency bins share one gain factor, the GMM modeling in each individual frequency bin is not easy to extend to GSMM. The alternative way, described in (4.32), however, can be extended easily since it models the probability over frequency bins.
This project only deals with DFT and DCT domain. Wavelet domain is also an option. Or even adaptive basis function can be used, such as basis in [35].

When source separation is embodied in feedback cancellation, strategy I uses a fixed step size. It can also be generalized as a time-varying step size. For example, as it is known from the simulation, the weighted adaptive filter doesn’t work well for signals with very different energy. Therefore, the step size can be normalized by the energy difference between the incoming signal and feedback signal estimated from feedback canceller, i.e., in inverse proportion to the energy difference. Another new idea is to stop or slow down the adaptation of the parameters in source separation (e.g., GMM) when feedback canceller by other means, such as whistle detection, is judged to be misbehaving. In this way, source separation will not be misled by feedback canceller anymore and the help from source separation can be lengthened. At the same time, the step size $\alpha$ could be proportional to the absolute difference between estimation from the feedback canceller and source separation to strengthen the helping force. When the feedback canceller recovers from bad behavior, the difference is big due to the “lag” effect of source separation. In this case, the step size could be inverse proportional to the difference, so that the adaptation of feedback canceller is not hindered by the lag source separation Thus, other means to judge the behavior of feedback canceller might be necessary.

A new question that might be worth investigating is that: if source separation can be developed into an “adaptive reference filter”? A disadvantage of constrained feedback canceller is the invariance of reference filter that limits the adaptation to new environments. With an adaptive reference filter, the merits of constrained feedback canceller can be retained while the disadvantage is minimized.


Simulations results


