Electronics Letters is published fortnightly except in December (2 issues in March)

2001 annual subscription prices:
print or online £199.00/US$315.00, combined print/online £249.00/US$391.00. Single copy £49.00/US$65.00

Air mail service direct to subscribers in an additional £25.00/US$40.00

All subscription inquiries and orders should be sent to IEEE Publications Sales, PO Box 98, Stevenage SG1 5AS, United Kingdom.
Telephone: +44 (0)1438 747293, Fax: +44 (0)1438 745799, E-mail: sales@ieee.org

All other correspondence should be sent to the IEEE Publishing Department, Michael Pindar House, Six Hills Way, Stevenage SG1 4AB, United Kingdom. Telephone: +44 (0)1438 767298, Fax: +44 (0)1438 767375, E-mail: letters@ieee.org

Scope and guides to authors
Electronics Letters emphasises electronic science and engineering, telecommunications, optoelectronics and optical communication. Electronics Letters specialises in the rapid publication of short international research papers at the cutting edge of electronics technology. Contributions should be in accordance with the Guide to Authors, found on the inside back cover or at http://www.ieee.org/PublicationsStandards/Proceedings/journals/

Electronics Letters Online
Electronics Letters is available online through the IEEE Online Journals service on the IEEE's World Wide Web server. Articles are stored in PDF format and may be viewed and printed using the Adobe AcroRd Reader. Postscript is also available. Please note: tables of contents and abstracts are provided for each journal issue, cross-journal searching of bibliographic records (including authors, titles and abstracts) and the ability to print articles locally. The issues are added in advance of print publication.

For further information see http://iee.org/, email letters@ieee.org or phone +1 732 562 5040.

Copyright and copying
This publication is copyright under the Berne Convention and the Universal Copyright Convention. All rights reserved. Apart from any copying under the UK Copyright, Designs and Patents Act 1988, Part I, Section 31, whereby a single copy of an article may be supplied, under certain conditions, for the purpose of research or private study, by a library of a class prescribed by The Copyright, Libraries and Archives Board (Copies of Copyright Material) Regulations 1988 S.I. 1988/1022, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means without the prior permission of the copyright owners. Permission is, however, not required to copy abstracts of individual contributions on condition that full reference to the source is given.

Single copies may be made for the purpose of research or private study.

Authorisation to photocopy items for internal or personal use, or for the internal or personal use of specific clients, is granted by the IEEE for libraries and other users registered with the Copyright Clearance Center Transactional Reporting Service, provided that the base fee of $2.00 per copy is paid directly to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA, 00026 (508) 750 8400. This authorisation does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works or for resale.

Multiple copying of the content of this publication without permission is always illegal.

The IEEE is not a body responsible for the opinions expressed by individual authors in Electronics Letters.

The IEEE is a member of the Association of Learned & Professional Society Publishers.

Chief Executive of the IEEE: AII Roberts Phel Org IEEE FrontP
Publisher: Balina Mather-Source
Managing Editor: Aibinika Kazakiewicz
Executive Editor: Helen Dyball
Editorial Assistants: Carol Marriott, Paul Bowler, Natasha Deverson and Wendy Phillips
Honorary Editors: Prof. Pauli Bayvel FIEEE, Prof. Peter Claremont CBE, FIEEE FRS, Prof. Chris Toumazou and Prof. Ian White

Typeset by: Typetec Composition, 12 The Malverns, Cheery Orchard Lane, Salisbury SP2 8JG, United Kingdom
Printed by: A. McKay & Co Limited, Langwood Drive, Forest Farm, Cardiff CF4 7SR, United Kingdom

© 2003: The Institution of Electrical Engineers
feature selection, the Hamming window size is 32 ms and silence and unvoiced segments are discarded based on an energy
and only the first session (24 sentences) is used for training. A more general learning method over various conditional probability
distributions has been described in [3].

Testing: The aim in the testing procedure of speaker identification is to determine the right person given an observation, i.e., max
p(M_j|O), i = 1, . . . , N, where M_j is the model of speaker i. It means we have to calculate the posterior probability, p(M_j|O), i = 1, . . . , N.
According to the Bayes rule, p(M_j|O) = p(O|M_j)p(M_j)/p(O). Since no knowledge about the prior probability p(O) is known and probability
p(M) is the same for all the models, we use p(O|M_j) in substitute of p(M_j|O) for simplicity. It can be achieved by computing the joint
probability using (9).

Experiments: In our experiments, we defined the topology of the DBNs as shown in Fig. 1, which is unrolled for first two slices. q_{it},
i = 1, 2, 3, j = 1, 2, . . . , T are hidden nodes and have discrete values, q_{it}, i = 1, 2, 3, j = 1, 2, . . . , T can be observed and satisfy Gaussian
distributions, here T is the length of time slices.

For computational reasons, a subset of the YONO database including the first 30 speakers (speaker ID from 101 to 132, excluding 123 and
129 since no data of these speakers exist in YONO corpus) is used, and only the first session (24 sentences) was used for training. In the feature
selection, the Hamming window size is 32 ms (256 samples), and silence and unvoiced segments are discarded based on an energy
threshold. The feature vector is composed by 16 MFCC. All the verify
sentences (40 per user) are used in the test procedure. We compare the recognition rates over different numbers of speakers of DBNs to
other classical methods such as Gaussian mixture model (GMM),
continuous hidden Markov model (CHNM), single Gaussian and
topic quantisation (VQ) in Table 1. It can be seen that the DBNs-based method has achieved encouraging results.

Table 1: Recognition rate results under various methods over different number of speakers

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of speakers</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st 10% (%)</td>
<td>20%</td>
<td>30%</td>
<td></td>
</tr>
<tr>
<td>DBNs</td>
<td>100</td>
<td>97.250</td>
<td>96.667</td>
<td></td>
</tr>
<tr>
<td>GMM</td>
<td>99.500</td>
<td>95.125</td>
<td>94.750</td>
<td></td>
</tr>
<tr>
<td>CHMM</td>
<td>99.250</td>
<td>94.125</td>
<td>94.416</td>
<td></td>
</tr>
<tr>
<td>Single Gaussian</td>
<td>90.000</td>
<td>94.000</td>
<td>92.667</td>
<td></td>
</tr>
<tr>
<td>VQ</td>
<td>98.300</td>
<td>94.625</td>
<td>94.083</td>
<td></td>
</tr>
</tbody>
</table>

Conclusion: Dynamic Bayesian networks are expressive models for stochastic processing. We have presented a framework using
dynamic Bayesian networks for speaker identification and have achieved considerable results in the experiments on a subset of
YONO corpus.

Acknowledgments: We wish to acknowledge support for this work by the National High Technology Research & Development Programme
(863) of the People's Republic of China (No. 2001AAA180) and the Zhejiang Provincial Natural Science Foundation for Young Scientists
of the People's Republic of China (No. RC01058).

© IE 2003
Electronics Letters Online No: 20030184
DOI: 10.1049/el:20030184
LiFeng Sang, Yingchun Yang, Zhaochui Wu and WanFeng Zhang
(Department of Computer Science & Engineering, Zhejiang University, Hangzhou, 310027, People's Republic of China)
E-mail: lfsang@cc.zju.edu.cn

Real-time room acoustic response simulation by IIR adaptive filter

G. Costantini and A. Uncini

A new IIR adaptive filter for real-time, room acoustic response
simulation is proposed, the structure of which derives from Jot's model of an artificial reverberator. The simultaneous perturbation
stochastic approximation (SPSA) algorithm is used to set parameter
values. Results show good similarity between the desired and artificial
response.

Introduction: It is common knowledge that different types of music
require different acoustical characteristics, related to their particular
requirements. Every closed space induces some reflections of the
signal generated inside and, together with other phenomena, have
influence on the auditory feeling of a listener. For example, listening
to a sacred music choir requires a very reverberant place, that simulates the complex of vocal sounds, and works like a powerful
case of resonance; a piano concert, however, requires a drier ambient, that allows careful distinction of every single note. Therefore, there exists the problem of rendering a listening place adaptable to different types of music, artificially reproducing acoustical characteristics of a particular
ambient. Our research objective is to obtain an accurate reproduction of an acoustical impulse response (IR) of a generic room through a
real-time musical signal processing, introducing similar reverberation and spectral characteristics of a target ambient.

Reverberation: The reverberation phenomenon consists in a persist-
ence of sound gradually attenuated during a certain temporal interval
after the sound source has stopped. This is due to the multiple wall
reflection of acoustic spherical waves in the listening room, which
reflects them with an absorption coefficient α, dependent on the
constitutive materials [1]. Reflected waves can be divided into two
classes. First, there are the early reflections, i.e., waves reflected only
once before reaching the listener; this part is very important to create
in the listener's brain the idea of the room spatial dimensions. Then,
many other wave-rays are reflected more times by the walls before
reaching the listener, thus creating a very dense whole of echoes that
arrive at the listener in random times and constitute the reverberating
tail.

To quantify the reverberation entity in a room, the reverberation time
T_60 is defined as the time for the impulse response decaying from 5 to

330 ELECTRONICS LETTERS 6th February 2003 Vol. 39 No. 3
Artificial reverberation: Schroeder was the pioneer who first attempted to make digital reverberation. The first prototype he tried, called comb filter (or plain reverberator), consisted of a single delay line of \( m \) samples with a feedback loop containing an attenuation gain \( g \). To simulate the frequency selective absorption of air and walls, Schroeder modified the plain reverberator, inserting a lowpass filter in the feedback loop, in order to increase the time echoes density and to decrease the metallic effect produced by comb filters, he cascaded multiple all-pass filters. Schroeder's structure is composed of four parallel comb filters, followed by two cascaded all-pass ones [2]. The Schroeder frequency response results provide a good quality, but it is rather "anonymous".

An evolution of Schroeder's model was developed by Moorer [3]. He observed that Schroeder's model can well modelise only the reverberation tail, while a room is characterised by its early reflections, that have a quite different: we seek to obtain a general purpose structure that can efficaciously simulate a desired IR must have the properties of:

- the reverberated signal sounds well, i.e. it comes out pleasant for the subjective hearing;
- the dry signal filtered by the identified reverberator is as similar as possible to the signal really produced inside the room, the IR of which is measured;
- the structure is purposed for any real-time implementation.

The structure chosen for the artificial reverberator is shown in Fig. 1. It is a generalisation of the filtering structure developed by Jot in 1992 [4].

For the \( G \) matrix we chose \( g_i = 0 \) (\( i = 1, \ldots, 6 \)), so every comb is feedback on all the others, except on itself. The \( b_i \) and \( c \) coefficients are employed for scaling every comb and all-pass contribution, thus increasing the elasticity of the filter. Thus the employed structure is characterised by the following parameters:

- \( g, g_2, D_i \) for the six comb filters (18 coefficients);
- the \( G \) matrix (36 coefficients);
- \( b_i \) and \( c \) (7 coefficients);
- \( g \) and \( R \) for the two all-pass filters (4 coefficients);

for a total number of 65 free coefficients. We can indicate the transfer function of the whole structure with \( F(z, \omega) \), where \( \omega \) indicates the free coefficients vector.

Identification procedure: The identification procedure requires finding the values of the 65 coefficients that make the IR of the filter as similar as possible to the desired IR. It is impossible to accomplish this procedure by setting the parameters by hand; we have to employ, on the contrary, an automatic procedure, in which the coefficients of the filter are iteratively corrected by a suitable adaptation algorithm, based on the difference \( e(n) \) between the desired response \( d(n) \) and the filter output response \( d(n) \), to make the error \( e(n) \) as low as possible.

The optimisation algorithm used for the reverberator filter identification is the simultaneous perturbation stochastic approximation (SPSA); it was developed by Spall in 1992 [5]. Let us consider the problem of minimising a scalar differentiable loss function \( L(W) \), where \( W \) is the \( p \)-dimensional vector of parameters: the optimisation problem can be translated into finding the minimising \( W^* \) such that \( \partial L/W = 0 \).

The choice of the loss function \( L(\cdot) \) represents the crucial point of the entire identification procedure: the final result depends on the capacity of the loss function to express the filter reverberation quality, also regarding the sentence 'it well sounds'. \( L(\cdot) \) must quantify the difference between the artificial response and the real response, as well as guaranteeing a good quality of artificial reverberation.

A structure that can efficaciously simulate a desired IR must have the properties of:

- **Elasticity**: the filter must have quite a number of free parameters and must allow free variations of its characteristics;
- **Robustness**: the filter must be stable and able to produce every kind of IR, with different \( T_{60} \);
- **Enough complexity**: the structure must generate an high echo density.

A structure that satisfies these properties is an "additive filter": the novelty that Jot introduced, compared to the Moorer model, is the presence, in the IR part, of a feedback coefficients matrix \( G \), this being for two main reasons:

- remixing in time the echoes, the aim being to cancel the periodicity effect introduced by the comb filters and also by the FIR filter;
- increasing the number of free parameters for a better approximation of the real IR.

The loss function we propose in this Letter carries out:

(i) A comparison between the power \( p(t) \) of simulated IR and real IR, to re-create the envelope and the echo distribution; we windowed the temporal axis in \( N \) windows, calculated the power in each window and minimised the maximum power. It allows us to simulate rather well the temporal behaviour of the real IR, reproducing it with the same \( T_{60} \) and echo distribution.

(ii) A comparison between the real and the artificial frequency responses, minimising the mean square error, to obtain a similar frequency colouration between the real and artificial entire frequency response.

(iii) A further frequency test, minimising the maximum shifting, to reduce the defects introduced by the employed structure. We windowed...
the frequency axis in $M$ windows; then minimised the maximum error between the frequency responses in each window.

**Simulation results:** The adaptive filter described above has been tested by reproducing impulse responses artificially generated by an acoustical editing tool based on ray tracing and image-source methods. We set this tool to generate typical IR, assuming rectangular perpendicular rooms.

We report the results related to a small ambient, with a medium wall absorption coefficient $\alpha = 0.95$. Figs. 2a and b show the time response and the frequency response, respectively, of this room, assuming a sample frequency $f_s = 11025$ Hz. For the FIR filter, we chose the first IR 60 ms. At the end of the identification procedure, we obtained the filter responses shown in Figs. 2c and d. A good similarity between the desired and the artificial time and frequency response can be observed: the artificial reverber reached the same $T_{60}$ and the same harmonic content of the desired test room responses.

© IEE 2003

9 December 2002

**Electronics Letters**

DOI: 10.1049/el:20030185

G. Costantini (University of Rome 'Tor Vergata' — Department of Electronics Engineering, Via del Politecnico, 1-00133 Rome, Italy)

E-mail: costantini@uniroma2.it

A. Uncini (University of Rome 'La Sapienza' — INFO-COM Department, Via Eudoxiana, 18-00184 Roma, Italy)

**References**

1. SABINE, W.C.: ' Reverberation' in LINDSAY, R.B. (Ed.): 'Acoustics: historical and philosophical development' (Downer, Hutchinson, and Ross, Stroudsburg, PA, USA, 1972, first published 1900)

**Simplification of soft-bit speech decoding and application to MELP encoded speech**

Xiaobei Liu and Soo Ngee Koh

A simplified soft-bit speech decoding (SBSD) algorithm is proposed and is used in decoding the encoded speech parameters from the mixed excitation linear prediction (MELP) coder. Simulation results show that the simplified algorithm can reduce significantly the decoding complexity and memory requirement of SBSD at a slight penalty of a marginal drop in performance.

**Introduction:** Soft-bit speech decoding (SBSD), as a method which successfully uses the residual redundancy to achieve error concealment, can be applied to almost all speech coding algorithms with significant improvements in the quality of the decoded speech when the encoded bits are received with errors [1-3]. The main reason that deters the use of SBSD in real-time applications is the exorbitant complexity and memory requirement of the decoding algorithm, especially for parameters that are encoded by a large number of bits.

The complexity issue is only considered in [2] in which the authors propose to employ lower dimensional vector quantisers (VQs). However, this approach requires the redesign of codebooks at the encoder end. In this Letter, a simple but effective simplified SBSD without modifying the encoder is proposed and is applied to mixed excitation linear prediction (MELP) encoded speech parameters [4, 5]. Simulation results show that while our proposed method reduces significantly the decoding complexity and memory requirement, it only costs a slight degradation in the quality of the decoded speech.

**Complexity of SBSD:** The SBSD algorithm can be divided into two steps. First, the decoder uses the received sequence $\hat{x}(i) = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N\}$ to compute the a posteriori probability (APP) of each of the possible transmitted bit combination $i = 1, \ldots, 2^M-1$. The resulting equation is:

$$P(T_i | \hat{x}_N) = C \cdot P(T_i | \hat{x}_{N-1}) \cdot \prod_{i=0}^{N-2} P(T_i | \hat{x}_{i+1})$$

where $C$ is the normalisation constant. Secondly, the APPs are used for MMSE estimation of the reconstructed parameter $T_i$ as shown in (2):

$$\hat{T}_i = \sum_{j=0}^{2^M-1} T_i \cdot P(T_i | \hat{x}_N)$$

From the above, we can see that the total complexity for calculating each parameter is about $2^M$ for both multiplication and addition. The memory required is $2^M$ also for storing the a priori index transition probabilities $P(T_i | T_{i-1})$.

**Simplified SBSD:** Suppose $T_0$ is the combination of the first $N$ significant bits of $\hat{x}_N$ which means $\hat{x}_N = \{\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_{N-1}\}$, $N \leq M$. To reduce the memory requirement of SBSD, we propose to divide the $2^M$ parameter indices into $2^N$ groups and each group includes $2^{M-N}$ of the same $\hat{x}_N$. The transition probabilities between groups are used instead of the individual index transition probabilities. The transition probability between each two groups is calculated as

$$P(T_i | T_{i-1}) = \frac{\sum_{j=0}^{2^M-1} T_i \cdot P(T_i | \hat{x}_N)}{\sum_{j=0}^{2^M-1} P(T_i | \hat{x}_N)}$$

Equation (1) now becomes

$$P(T_i | \hat{x}_{N}) = C \cdot P(T_i | \hat{x}_{N-1}) \cdot \prod_{i=0}^{N-2} P(T_i | \hat{x}_{i+1})$$

where

$$P(T_{i-1} | T_{i+1}) = \frac{\sum_{j=0}^{2^M-1} T_i \cdot P(T_i | \hat{x}_N)}{\sum_{j=0}^{2^M-1} P(T_j | \hat{x}_{N})}$$

From the above, it can be found that the memory requirement is reduced to $2^N$ and the numbers of multiplications and additions are $2^{M-N}$ and $2^M$, respectively. To further reduce the number of calculations, we propose to ignore the 'index transmission probabilities' $P(T_i | T_{i-1})$ below a certain value $T$ since they contribute negligibly to the reconstructed parameters. Suppose $R$ out of the $2^N$ index transmission probabilities remain after ignoring the others, the numbers of multiplications and additions then become less than $R \cdot 2^N$ and $R \cdot 2^M$, respectively.

**Results and discussion:** To test the performance of the proposed procedure, the simplified SBSD is first applied to the MELP encoded pitch parameter [6] which is quantised to 7 bits and requires 16 384 arithmetic operations and words. Table 1 shows the parameter SNR (PSNR) of pitch at $F_s/N_s = 0$ dB obtained by the simplified SBSD for different values of $T$ and $R$. It should be noted that the PSNR of pitch with the unmodified SBSD is 16.6 dB and its value without the use of SBSD is only 9.77 dB.

From Table 1, it can be observed that the performance improves with increasing value of $T$ and decreasing value of $R$. When $N>4$ or $T < 10^{-6}$, the amount of improvement becomes marginal whereas the number of arithmetic operations increases rapidly. When $T=10^{-6}$, there is even a slight decrease in the PSNR, which can be explained by the index transmission probabilities below $10^{-6}$ being very unlikely to be the correct transmitted indices and taking them into account in the reconstruction of the parameters will only degrade the decoder performance.

332 ELECTRONICS LETTERS 6th February 2003 Vol. 39 No. 3