IEEE GLOBECOM 1996

LONDON

November 18 - 22, 1996

GLOBECOM 96 London

Communications: The Key to Global Prosperity

GLOBAL INTERNET'96

Conference Record

Sponsored by the IEEE Communications Society, the IEE and the UKRI Communications Chapter

IEEE 1996 Global Telecommunications Conference

IEEE catalog number: 96CH35942
ISBN numbers:
softbound edition: 0-7803-3336-5
casebound edition: 0-7803-3337-3
microfiche edition: 0-7803-3338-1
Library of Congress number: 87-640337

Additional Proceedings may be ordered from:

IEEE Service Center
Publications Sales Department
445 Hoes Lane, P.O. Box 1331
Piscataway, NJ 08855-1331, USA

A New Technique for Audio Packet Loss Concealment

H. Sanneck*, A. Stenger, K. Ben Younes, B. Girod
Telecommunications Institute, University of Erlangen-Nuremberg
D-91058 Erlangen, Germany
sanneck@fokus.gmd.de, {stenger, younes, girod}@nt.e-technik.uni-erlangen.de

Abstract

We present a new error concealment technique for audio transmission over heterogeneous packet switched networks based on time-scale modification of correctly received packets. An appropriate time-scale modification algorithm, WSOLA ("Waveform Similarity Overlap-Add"), is used and its parameters are optimized for scaling short audio segments. Particular attention is paid to the additional delay introduced by the new technique. For subjective listening tests, packet loss is simulated at error rates of 20% and 33% and the new technique is compared to previous proposals by category and component judgment of speech quality. The test results show that typical disturbance components of other techniques can be avoided and overall sound quality is higher.

1 Introduction

Today's packet switched networks are more and more used not only for bulk data transfer, but for audio and video transmission. An example is the MBone overlay network in the Internet. However, packet switched networks typically offer only a "best effort" service, which does not make any commitment about a required minimum bit-rate or a maximum delay allowed. Consequently, when the network gets congested, real-time packets may arrive too late at the receiver or may be dropped at routers due to buffer overflow. In the case of the transmission of waveform-coded audio, packet loss causes signal drop-outs which are very annoying for the listener.

To alleviate this problem, different techniques have been proposed, which include modifications of the receiver *and* the transmitter ([1] - [5]). We have focused instead on receiver-only methods ([6] - [8]), as they do not introduce additional processing and data overhead at the transmitter and are well suited for heterogeneous multicast environments. This means that transmitters may use different audio tools than the receivers, and receivers can mitigate packet loss according to their specific quality requirements.

Section 2 gives a brief overview over common receiver-only algorithms. In section 3.1 we present a new concealment method based on a fast time-domain algorithm ([9]). The time-scale of a signal segment immediately following the gap caused by missing packets is modified such that this gap is covered by a *"stretched" version of the segment* (Fig. 1).

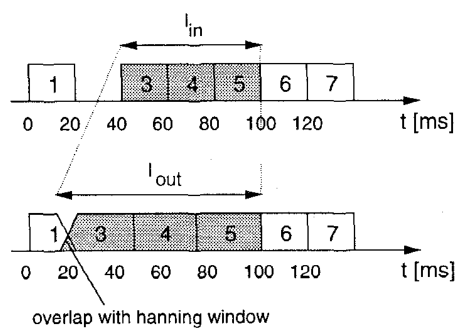

Figure 1: Error concealment using time-scale modification.

As the original algorithm is designed for signal segments of long duration, its parameters have to be chosen carefully to match the requirements of short signal segments (the contents of only a few audio packets) and thus provide for minimum extra delay. The selection of parameters is derived in section 3.2, and in section 3.3 the extra delay and additional computational effort introduced by the new technique are discussed. The results of listening tests are presented in section 4.

*now with the Research Institute for Open Communication Systems GMD FOKUS, Hardenbergplatz 2, D-10623 Berlin, Germany

2 Audio Packet Loss Recovery through Waveform Substitution

Previous proposals for receiver-only concealment of lost audio packets substitute the missing signal segment by *repeating a prior segment*. The *"Pattern Matching"* ([6], [7]) technique repeats a correctly received signal segment, of which maximum similarity with the lost segment is assumed. This is accomplished by matching a sample pattern immediately preceding the gap to series of samples received earlier. As entire signal segments of at least one packet duration are completely repeated, this may cause echoing sounds.

Echoes can be avoided by *"Pitch Waveform Replication"* ([6], [7], [10]), where only one pitch period found in the most recently received packet is repeated throughout the missing packet. An extension of this technique, called *"Phase Matching"*, provides for synchronization on both edges of the substitute, reducing a clicking distortion caused by the other two methods ([8]). In the latter two cases the multiple repetition of the same small signal segment can lead to tinny sounds.

All of these techniques have in common that, with an increased length of the lost segment, the perceived quality deteriorates severely. We believe that this decrease in quality is to a large extent due to the specific distortions introduced by the different concealment techniques. Therefore, in the following sections a new method, based on time-scale modification, is proposed.

3 Description of the new method

3.1 Time-scale Modification with the WSOLA Algorithm

An appropriate algorithm must perform time-scale modification in real-time, and it may not change the pitch frequency to preserve natural sound. The "Waveform Similarity Overlap-Add" (WSOLA) algorithm presented in [9] meets these requirements.

As shown in Fig. 2, time-scale expansion is achieved by extracting segments $\mathbf{A}_\nu$ from the input signal at time instances $t_{x\nu}$ and superimposing them in the output signal at larger spaced time instances $t_{y\nu}$. The time instances $t_{x\nu}$ are selected from a tolerance interval around $t_{x\nu-1} + T_x$, where the input step size T_x controls the time-scaling factor. To achieve a synchronized overlap of the $\mathbf{A}_\nu$ in the output signal, the candidate for $\mathbf{A}_\nu$ is chosen from the tolerance interval, whose leftmost L' samples resemble as close as possible $\mathbf{A}'_\nu$, where L' denotes the length of $\mathbf{A}'_\nu$. A computationally efficient similarity measure, the cross-correlation coefficient

$$c_c(\nu,\delta) = \sum_{k=0}^{L'-1} x(t_{x\nu-1} + T_x + k + \delta) \cdot x(t_{x\nu-1} + T_y + k)$$

is used ([9]). Then the desired time instance is

$$t_{x\nu} = t_{x\nu-1} + T_x + \delta^*, \qquad \delta^* = \text{argmax}_\delta\{c_c(\nu,\delta)\}.$$

The output step size is fixed to $T_y = \frac{1}{2}L$ (L being the length of the $\mathbf{A}_\nu$) and a fixed length hanning window is employed for overlap.

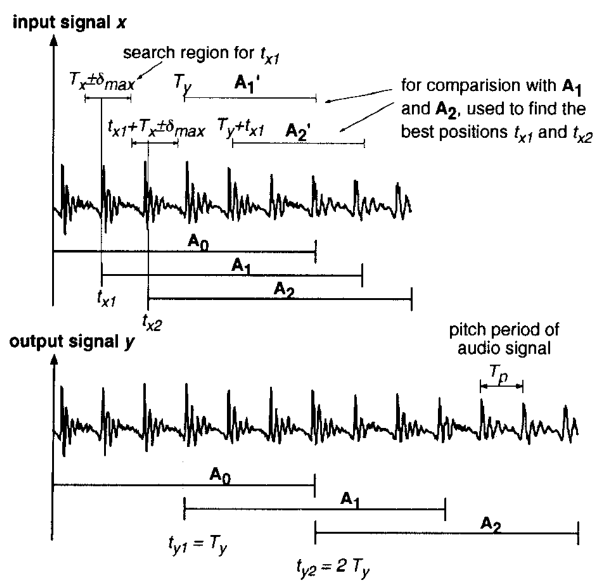

Figure 2: Enhanced speed version of the WSOLA algorithm.

3.2 Selection of Parameters

Fig. 1 shows the principle of error concealment using time-scale modification of correctly received packets. Distortions due to discontinuities at the boundary of the substituted and the next received packet are reduced by overlapping them in the range of 2 ms using hanning windows.

As the WSOLA algorithm was originally developed for time-scale modification of long speech sequences, the parameters T_x, δ, T_y and the lengths of $\mathbf{A}_\nu$ and $\mathbf{A}'_\nu$

must be chosen carefully to preserve the good sound quality when WSOLA is applied to short signal segments as shown in Fig. 1.

$\mathbf{A}'_\nu$ must include at least one pitch period, so that the correct synchronization can be found ($L' \geq T_{p,max}$, where $T_{p,max}$ denotes the maximum pitch period of a speech signal). In contrast to [9], we chose the maximum possible interval $[-T_x, T_y - T_x]$ for the parameter δ. The length of this interval must as well include at least one pitch period ($T_y = L/2 \geq T_{p,max}$). We obtain $L' \geq L/2$ for the relation of the lengths of $\mathbf{A}_\nu$ and $\mathbf{A}'_\nu$. In our experiments we set $L' = L/2$ ([9]: $L' = L$), which results in an enhanced speed version (Fig. 2) with no perceivable deterioration of the output signal.

In addition, there is the following relation between the length of the input series of samples l_{in} and the estimate l'_{in} of the number of samples used by the algorithm,

$$l'_{in} \approx (N-1)T_x + L \overset{!}{\leq} l_{in}$$

where N is the number of extracted segments ($0 \leq \nu < N$). Another constraint exists between the number of samples l_{out} needed to replace the missing speech segment and the length of the output series of samples l'_{out} received from the algorithm:

$$l'_{out} = (N-1)T_y + L = (N+1)\frac{L}{2} \overset{!}{\geq} l_{out}$$

For long speech sequences we have $L \ll l_{in}$, so the time-scale ratio $\beta = T_y/T_x$ is approximately equal to the expansion ratio $\alpha = l_{out}/l_{in}$. However, for very small input lengths the above condition doesn't hold, which results in $\beta \gg \alpha$. This implies extreme time expansion and poor output quality. Thus it is necessary to chose α as a compromise of output quality and additional delay introduced. Additionally, an optimization algorithm for the parameter $\varepsilon = l'_{out} - l_{out}$ might be employed to minimize loss of information. A further suggestion for the choice of these parameters can be found in [11].

3.3 Discussion of Delay and Computational Complexity

While other waveform substitution techniques *repeat* information from correct packets, our new scheme also *modifies* these packets themselves. Thus the computational effort to execute the algorithm is higher. Additionally, it is necessary not only to keep a copy of a certain number of packets, but to withhold these packets from the playout buffer.

Considering the parameter relations described in the previous paragraph, we chose to conceal one lost packet (160 samples PCM 8 kHz (G.711) $\rightarrow$ 20ms audio) with three following packets. Four packets are needed to recover from the loss of two consecutive packets. In Fig. 1, packets 1 - 5 have to be queued for WSOLA and packets 1 - 3 for the other techniques, if overlap-add with correct neighbor packets is performed (for the other algorithms, copies of packets preceding packet 1 have to be stored dependent on the length of the search area). In communication systems that artificially introduce audio delay, such as video conferencing systems ([12]), the extra delay is acceptable. Furthermore, audio tools like *vat* and *NeVoT* include delay adaptation mechanisms ([13]), which add some delay to cope with packet delay jitter.

In the above example, twice the number of additions and four times the number of multiplications have to be performed, compared to the other algorithms (see [14] for a detailed analysis). Taking into account the absolute number of operations ($\approx 8 \cdot 10^4$ multiplications and $\approx 4 \cdot 10^4$ additions for the example) and the execution time usually available (because of delay adaptation) as well as the computing power of today's average hardware, the computational effort can still be considered as low.

4 Experimental Results

4.1 Test Environment

Four speech signals with different pitch frequencies (two male and two female speakers), sampled at 8 kHz, were used. The new Time-scale Modification technique (TM) was compared to Silence Substitution (S), Pattern Matching (PM) and Pitch Waveform Replication (PWR) using 40 test conditions of 10 seconds each. Thirteen non-expert listeners were asked to judge overall quality on a five-category (Mean Opinion Score) scale, comparable to test schemes used in [1], [2], [7] and [8]. Additionally the presence of the disturbance components "tinny, metal", "interrupted, clicking" and "echoing, reverberating" for each condition was judged.

We modelled single packet loss by suppressing one packet within five ("1 of 5") and double packet loss by

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase
Smarter legal research.