AN ANALYSIS OF INTERLEAVERS FOR ROBUST SPEECH RECOGNITION IN BURST-LIKE PACKET LOSS

A.B. James, B.P. Milner

School of Computing Sciences, University of East Anglia, Norwich, U.K.
{a.james, b.milner@uea.ac.uk}

ABSTRACT

An analysis into the effect of packet loss shows that a speech recogniser is able to tolerate large percentages of packet loss provided that burst lengths are relatively small. This leads to the analysis of three types of interleaver for distributing long bursts of packet loss into a series of shorter bursts. Cubic interpolation is then used to estimate lost feature vectors. Experimental results are presented for a range of channel conditions and demonstrate that interleaving offers significant increases in recognition accuracy under burst-like packet loss. Of the interleavers tested, decorrelated interleaving gives superior recognition performance and has the lowest delay. For example at a packet loss rate of 50% and average burst length 20 packets (40 vectors or 400ms) performance is increased from 49.6% with no compensation to 86% with interleaving and cubic interpolation.

1. INTRODUCTION

The growth of mobile and handheld devices for speech communication has resulted in distributed speech recognition (DSR) systems being developed. The European Telecommunication Standards Institute (ETSI) Aurora DSR standard [1] offers good robustness to noise by replacing the low bit-rate speech codec on the terminal device with the static MFCC feature extraction component of the speech recogniser. Figure 1 shows an overview of a typical DSR system, along with the proposals outlined in this work.

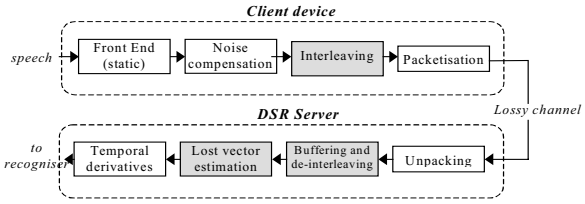

Figure 1: Architecture of an interleaved DSR system.

DSR systems often transmit speech data in the form of packets (or frames) across networks that do not guarantee reliable delivery. If these packets become lost, or too many bits are corrupted so that bit level forward error correction cannot correct the frame, then portions of the feature vector stream become lost.

Early work on packet loss compensation for DSR considered splicing the feature vectors stream together in loss periods [2] or repetition of correctly received vectors to compensate for lost vectors [3]. Alternative schemes have used interpolation to estimate lost packets [4] and also the provision of error correction bits to minimise packet loss [5]. These schemes have varying degrees of success and work reasonably well for short duration bursts of loss but degrade as burst lengths increase.

The aim of this work is to improve speech recognition robustness in burst-like packet loss by comparing the effect of three types of interleaver. An analysis into the effect of both the percentage of packets lost and average burst length on speech recognition accuracy is made in section 2. Based on this analysis section 3 considers three types of interleaver for dispersing bursts of packet loss. Cubic interpolation is described in section 4 for estimating lost vectors. Section 5 measures the effectiveness of the interleavers in terms of recognition accuracy and delay. A conclusion is made in section 6.

2. THE EFFECT OF PACKET LOSS ON DSR

The conditions that cause packet loss on both mobile and IP networks often have sufficient duration to effect several concurrent packets and therefore result in burst-like packet loss. Two metrics are considered for characterising such a channel condition; namely the packet loss rate, α, and the average burst length, β . Figure 2 shows how these two characteristics affect recognition accuracy for packet loss rates from 10% to 50% and average burst lengths from 1 to 20 vectors – see section 5 for experimental details. The scheme in figure 2a employs no packet loss compensation with the result that accuracy is largely governed by the packet loss rate, α, whilst the average burst length, β, has far less effect. It is interesting to observe that as the burst length increases, the accuracy converges to:

baseline accuracy × (1 – proportion of vectors lost)

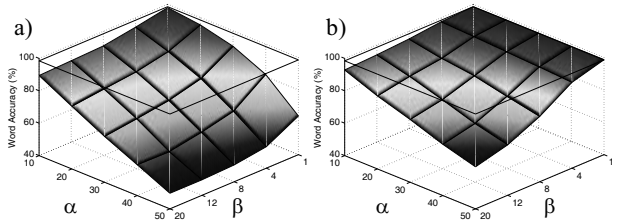

Figure 2: Word accuracy against varying channel condition with: a) no compensation , b) interpolation.

The scheme in figure 2b uses interpolation to estimate the value of lost vectors. In this scheme the overall loss rate, α, has less effect on accuracy than the average burst length, β. This is because interpolation is more effective at correcting short duration bursts of loss. As burst lengths increase it becomes

more difficult to accurately estimate missing vectors and hence accuracy falls.

These results show that when attempting to estimate lost vectors it is not the proportion of vectors lost that is significant, but rather the average burst length. Indeed, baseline accuracy of 98.6% can be maintained even at a loss rate of 50% providing the average burst length is short. Thus, for DSR, it is more important to reduce the average burst length of lost vectors rather than to reduce the overall packet loss rate through channel coding schemes. An effective technique for reducing burst lengths is to interleave the feature vectors before packetisation.

3. INTERLEAVING

Interleaving is applied to DSR on the terminal device and serves to permute the order in which feature vectors are packetised such that in the event of packet loss, consecutive feature vectors are not lost. Formally, for a sequence of feature vectors, $\boldsymbol{X}$, where,

$$\boldsymbol{X} = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N-1}\} \quad (1)$$

interleaving can be expressed as a permutation producing a re-ordered sequence, $\boldsymbol{X}'$, given as,

$$\boldsymbol{X}' = \{\boldsymbol{x}_{\pi(0)}, \boldsymbol{x}_{\pi(1)}, \boldsymbol{x}_{\pi(2)}, \ldots, \boldsymbol{x}_{\pi(N-1)}\} \quad (2)$$

The interleaving function, $\pi(i)$, gives the index of the vector to be output at the i^{th} time instance. Feature vectors are returned to their original order on the receiver side through de-interleaving which is given by the inverse function of π, i.e.,

$$\pi^{-1}(i) \quad where \quad \pi\left(\pi^{-1}(i)\right) = i \quad (3)$$

Conversely, the de-interleaving function $\pi^{-1}(i)$ gives the time instant that vector i is output. Figure 1 shows the location of the interleaver and de-interleaver in the DSR architecture.

The re-ordering made by the interleaving function means that feature vectors need to be buffered prior to transmission. For DSR applications this delay should be kept small. The interleaving delay, δ, is defined as the maximum delay that any vector experiences before being transmitted,

$$\delta = \max_i\left(\pi^{-1}(i) - i\right) \quad (4)$$

The ability of an interleaver to disperse bursts of loss is related to its spread. An interleaver has spread s if all pairs of vectors that are within s time instances of each other in the input sequence are separated by at least s time instances in the output sequence,

$$|x - y| \geq s \ \ whenever \ \ |\pi(x) - \pi(y)| < s \quad (5)$$

A burst of packet loss of length β will be totally distributed (i.e. no concurrent packets will be lost) by an interleaver with spread s if $s \geq \beta$. For the case $s < \beta$ the interleaver will not be able to fully distribute the burst, which will result in some consecutive packets being lost.

Both the spread and delay of an interleaver are functions of its degree, d. The degree of an interleaver relates to its buffer size but is considered differently for the various classes of interleaver. The remainder of this section considers three classes of interleaver in terms of their degree, and resulting spread and delay, for application to DSR.

3.1 Optimal spread block interleavers

A block interleaver of degree d operates by re-arranging the transmission order of a $d \times d$ block of input vectors. Two block interleavers, π_{block1} and π_{block2}, [6] are considered optimal in terms of maximising their spread for given degree, and are given,

$$\pi_{block1}(id + j) = (d - 1 - j)d + i \quad where \ 0 \leq i,j \leq d-1 \quad (6)$$

$$\pi_{block2}(id + j) = jd + (d - 1 - i) \quad where \ 0 \leq i,j \leq d-1 \quad (7)$$

It is interesting to observe that π_1 and π_2 form an invertible pair as $\pi_1 = \pi_2^{-1}$ *and* $\pi_2 = \pi_1^{-1}$. The operation of these interleavers can be considered as a rotation of the $d \times d$ feature vector buffer either 90° clockwise or 90° anti-clockwise as shown in figure 3.

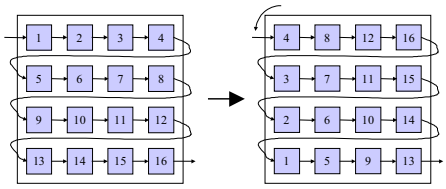

Figure 3: Rotation of buffer by 90° anti-clockwise.

The delay and spread of the two interleavers is related to their degree. From equations 6 (or 7), 4 and 5 the block interleaver delay, δ_{block}, and spread, s_{block}, are given as,

$$\delta_{block} = d^2 - d \quad \text{and} \quad s_{block} = d \quad (8)$$

3.2 Convolutional interleavers

Convolutional interleavers can be modelled as an arrangement of shift registers, each holding one feature vector [7]. In a convolutional interleaver of degree d, sequential input feature vector are divided amongst d sub-sequences. Each sub-sequence consists of a different number of connected shift registers and hence imposes a different delay to the feature vectors stored in it. A convolution interleaver of degree 4 is illustrated in figure 4.

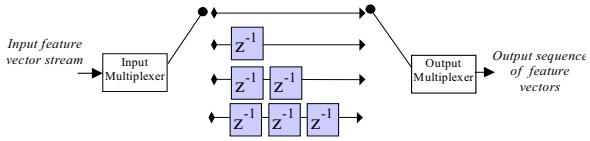

Figure 4: Convolutional interleaver of degree 4.

The interleaving function of a convolutional interleaver of degree d takes the form,

$$\pi_{conv}(i) = i - d(i \mod d) \quad (9)$$

The delay and spread of a convolutional interleaver are related to its degree. From equations 9, 4 and 5 the convolutional interleaving delay, δ_{conv}, and spread, s_{conv}, are given as,

$$\delta_{conv} = d^2 - d \quad \text{and} \quad s_{conv} = d - 1 \quad (10)$$

3.3 Decorrelated block interleavers

The previous interleavers aim to disperse burst-like loss by maximising spread according to equation 5. However an

alternative view of interleaving is that it is the process of decorrelating the order in which vectors are output in comparison to their input order. In this view, maximising the decorrelation will minimise the resulting average burst lengths.

A block interleaver, π, of degree d consists of a permutation sequence of length d^2. From this sequence a decorrelation measurement, D_π, can be made,

$$D_\pi = \sum_{i=1}^{d^2}\sum_{j=1}^{d^2}\frac{|\pi(i)-\pi(j)|}{|i-j|} \qquad (11)$$

It can be shown that the ability of an interleaver to distribute bursts of packet loss is directly related to this decorrelation value. This is demonstrated in an experiment whereby a set of 1000 block interleavers, with random permutation sequences each of length 16, is generated – $\{\pi_1$ to $\pi_{1000}\}$. A channel is simulated with packet loss rate α=50% and average burst length β=4 with each packet transporting 2 vectors. Figure 5a shows the output average burst length as a function of decorrelation value for each interleaver. A speech recogniser is then applied to the resulting feature vectors (as described in section 5) and the digit accuracy shown against decorrelation value in figure 5b.

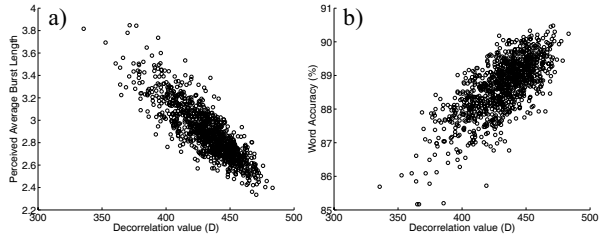

Figure 5: Decorrelation value against a) average burst length and b) word accuracy for 1000 random block interleavers.

The strong negative correlation shown in figure 5a indicates that interleavers with high decorrelation values are more effective at distributing bursts of packet loss than those with lower decorrelation values. Figure 5b shows that interleavers with high decorrelation values attain higher recognition accuracy due to the shorter duration bursts over which estimation must operate.

For a block interleaver of degree, d, the choice of permutation sequence to maximise the decorrelation value is not elementary. The number of possible permutation sequences of length d^2 is $d^2!$, hence a comprehensive state space search becomes impractical for higher degree interleavers. Heuristic search methods allow longer permutations to be created, but do not guarantee that results will be optimal. The decorrelated interleavers used in this work have been selected using a greedy local search [8], whereby movement in the state space is defined by the swapping of two elements in the permutation sequence. Once an suitable interleaver has been found its delay and spread can be found from equations 4 and 5.

4. ESTIMATION OF LOST VECTORS

The purpose of interleaving is to reduce the average burst length of the de-interleaved sequence such that estimation of the lost feature vectors is more effective. This work has considered a number of methods for estimating lost vectors and found that non-linear interpolation using cubic Hermite polynomials gives best estimates. The interpolation function for estimating the n^{th} lost vector in a burst of length β, starting at vector $b+1$ is

$$\hat{\mathbf{x}}_{b+n} = \mathbf{a}_0 + \left(\frac{n}{\beta+1}\right)\mathbf{a}_1 + \left(\frac{n}{\beta+1}\right)^2\mathbf{a}_2 + \left(\frac{n}{\beta+1}\right)^3\mathbf{a}_3 \qquad 1 \le n \le \beta \qquad (12)$$

The multivariate coefficients, $\{\mathbf{a}_0,.., \mathbf{a}_3\}$, need to be calculated so that vectors at the start and end of the loss follow a smooth trajectory with the first derivatives of the polynomial being continuous at the start and end of the loss [9]. These coefficients can be computed from the two vectors preceding and following the burst of loss, $\mathbf{x}_b$ and $\mathbf{x}_{b+\beta+1}$, and their first derivates, $\mathbf{x}'_b$ and $\mathbf{x}'_{b+\beta+1}$. Expressing the interpolation function in terms of Hermite basis functions gives the estimate of the n^{th} feature vector within the burst as

$$\begin{aligned}\hat{\mathbf{x}}_{b+n} = {} & \mathbf{x}_b\left(1-3t^2+2t^3\right)+\mathbf{x}_{b+\beta+1}\left(3t^2-2t^3\right) \\ & +\mathbf{x}'_b\left(t-2t^2+t^3\right)+\mathbf{x}'_{b+\beta+1}\left(t^3-t^2\right) \qquad 1 \le n \le \beta\end{aligned} \qquad (13)$$

where $t=n/(\beta+1)$, and derivates are approximated by $\mathbf{x}'_b = \beta\left(\mathbf{x}_b - \mathbf{x}_{b-1}\right)$ and $\mathbf{x}'_{b+\beta+1} = \beta\left(\mathbf{x}_{b+\beta+2} - \mathbf{x}_{b+\beta+1}\right)$. In practice it was found that rapid fluctuations in the feature vector stream resulted in large estimates of the derivative components causing the interpolation to overshoot. Improved performance was achieved by setting the derivative components to zero, leaving the interpolation function to comprise just the first two Hermite basis functions.

5. EXPERIMENTAL RESULTS

This section examines experimentally the trade-off between recognition accuracy and delay for the three classes of interleaver.

5.1 Recognition accuracy

The experimental results in this section examine the effect that interleaver degree has on recognition accuracy for a variety of simulated channels. The recognition task for these experiments is the Aurora connected digit database [1]. Digits are modelled using 16-state, 3-mode HMMs, trained from a set of 8440 digit strings. The test set comprises 4004 noise-free digits strings (13,159 digits in total) which gives baseline accuracy of 98.5% with 95% confidence error bands of +/- 0.76% at 75% accuracy and +/- 0.38% at 95% accuracy. As per the ETSI standard, two vectors are carried by each packet.

Four channels were simulated by a 3-state Markov chain [4] to give a mixture of network conditions in terms of the packet loss rate, α, and average burst length, β. Table 1 shows the parameters of these channels together with baseline recognition accuracy with no packet loss compensation.

Channel	Loss rate, α	Av. burst length, β	Baseline accuracy	
			No comp.	Cubic Int.
A	10%	4	91.19%	95.58%
B	10%	20	89.43%	90.55%
C	50%	4	49.56%	80.47%
D	50%	20	49.61%	56.34%

Table 1: Simulated channel conditions.

Shown in the final column is the recognition accuracy attained using the cubic interpolation of section 4, but with no interleaving. Experimental results, shown in figure 6, measure the effect of degree on recognition accuracy for the optimal spread block, convolutional and decorrelated interleavers.

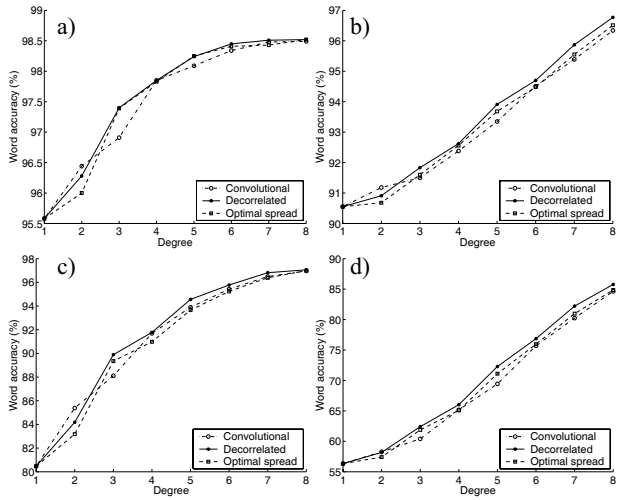

Figure 6: Recognition accuracy as a function of degree for various interleavers and channel conditions.

The figures show that interleaving feature vectors prior to transmission results in a significant increase in word accuracy, the magnitude of which is related to interleaver degree. Note that interleaving with degree 1 is equivalent to no interleaving, hence all graphs start with the accuracy specified for cubic interpolation in table 1. Figures 6a and c show the increase in accuracy levelling off as the spread of the interleavers becomes sufficient to fully distribute the bursts – this occurs at a degree of 8 for the average burst length of 4 packets (8 vectors). Increasing the degree beyond this point gives no further increase in recognition accuracy. For longer burst lengths shown in figures 6b and 6d the degree is not sufficient to fully distribute the bursts but does offer some gain in accuracy. To fully distribute these bursts a degree of at least 40 would be required. The figures also show that whilst all interleavers offer similar performance gains, the decorrelated interleaver generally results in slightly higher accuracy than the other interleavers.

5.2 Delay

The buffering necessary to permute the order in which vectors are transmitted introduces a delay that is related to the degree of the interleaver. Figure 7 illustrates this delay (measured in terms of feature vectors) as a function of degree for the three classes of interleaver described in section 3.

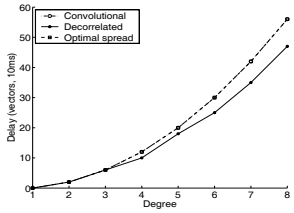

Figure 7: Interleaver delay as a function of degree.

The results show an exponential increase in delay as the degree is increased. Both the optimal block interleaver and convolutional interleaver have identical delays for a given degree, while the decorrelated interleaver has less delay, particularly at higher degrees. This illustrates the importance of selecting the correct degree such that word accuracy is maximised for minimal delay.

6. CONCLUSIONS

This work has shown that packet loss can have a severe effect on recognition accuracy. In particular the burst length is shown to be more detrimental to performance than the absolute proportion of lost packets. This suggests that improved recognition performance can be obtained by reducing average burst lengths through interleaving. Three types of interleaver have been considered for application to DSR and these were analysed in terms of their degree, which affects both the resulting spread and delay. Experiments showed that increasing the degree of the interleaver gives substantial increases in recognition performance, but also resulted in an exponential increase in delay. It is therefore important to match the design of the interleaver to the likely channel condition and maximum permissible delay. Of the three interleavers tested, the decorrelated interleaver is shown to give slightly superior performance in terms of higher recognition accuracy and lower delay for a given degree.

7. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the UK Engineering and Physical Sciences Research Council (EPSRC) in this work.

8. REFERENCES

[1] ESTI document - ES 201 108 – STQ: DSR – Front-end feature extraction algorithm; compression algorithm, 2000

[2] Kim, H.K. and Cox, R.V., “A bitstream-based front-end for wireless speech recognition on IS-136 communication system”, IEEE Trans. SAP, Vol. 9, No. 5, pp. 558-568, July, 2001

[3] Milner B.P., James A.B. “Analysis and compensation of packet loss in distributed speech recognition using interleaving”, Proc. Eurospeech, 2003

[4] Milner, B.P., “Robust speech recognition in burst-like packet loss”, Proc. ICASSP, 2001.

[5] Boulis, C., et al, “Graceful degradation of speech recognition performance over packet-erasure networks”, IEEE Trans. SAP, vol. 10, No. 8, pp. 580-590, November, 2002.

[6] Andrews K, Heegard C, Kozen D. “A theory of interleavers”, Technical report 97-1634, Computer Science Department, Cornell University, June 1997.

[7] Taub, H. and D.L. Schilling, “Principles of communication systems”, McGraw-Hill, 1986.

[8] Russell S, Norvig P., “Artificial Intelligence: A modern approach”, Prentice Hall, second edition, 2003.

[9] Vaseghi, S.V., “Advanced digital signal processing and noise reduction”, John-Wiley, second edition, 2000.