Low complexity near-optimal unit-selection algorithm for ultra low bit-rate speech coding based on N-best lattice and Viterbi search

V. Ramasubramanian, D. Harish

Siemens Corporate Technology - India, Bangalore, India
V.Ramasubramanian@siemens.com, D.Harish@siemens.com

Abstract

We propose a low complexity unit-selection algorithm for ultra low bit-rate speech coding based on a first-stage N-best pre-quantization lattice and a second-stage run-length constrained Viterbi search to efficiently approximate the complete search space of the fully-optimal unit-selection algorithm recently proposed by us. By this, the proposed low complexity algorithm continues to be near-optimal in terms of rate-distortion performance while having highly reduced complexity.

Index Terms: Speech coding, unit-selection, N-best search

1. Introduction

In a recent work [2], [3], [4] we proposed a unified and generalized framework for segment quantization of speech at ultra low bit-rates of 300 bits/sec based on unit-selection principle using a modified one-pass dynamic programming algorithm. While this one-pass DP algorithm offers the optimal rate-distortion performance, it suffers from a very high computational complexity of the order of \( O(M^2 T) \), where \( M \) is the number of frames in the continuous unit database (of the order of 8 – 13 \( \times 10^5 \) for a 17-bit unit-database) and \( T \) is the number of frames in the input test utterance being quantized (typically 50 for 1 sec of speech).

2. Proposed N-best lattice search

In this paper, we propose a N-best lattice search based two-stage unit-selection algorithm for reducing the complexity even while retaining a near-optimal performance of the fully optimal one-pass DP algorithm. This algorithm works as follows: i) Pre-quantize the input utterance using a variable-length segment quantizer (VLSQ) to yield a segmentation and corresponding VLSQ codeword labels; ii) Derive a N-best lattice of this pre-quantization using the N-best VLSQ segment labels of each segment; iii) For each segment in the pre-quantization, hypothesize a collection of units from the large continuous unit-database having the same labels as the union of all labels in the N-best lattice for that segment; iv) Perform a constrained Viterbi-decoding with concatenation costs (to favor run-length sequences) on these (N-best lattice based) unit-database units.

We bring out several important properties of this algorithm as well as its performances from Fig. 1 and Fig. 2: i) For \( N = 1 \), an earlier unit-selection algorithm [1] becomes a 1-best special case (blue-circle, Fig. 1) of the algorithm proposed here; however, the proposed N-best hypothesis generalizes to a larger search space and has significantly improved rate-distortion (R-D) for increasing \( N \) (dashed line curves).

ii) The proposed algorithm offers a progressive lowering of R-D curves (Fig. 1) towards left-bottom for increasing \( N \) (from 1 to 30), and even \( N = 30 \) reaches a performance that is ‘near-optimal’ to fully-optimal one-pass DP algorithm (the lowest and best R-D curve at the left-bottom - red-square, Fig. 1).

iii) This near-optimal rate-distortion trend is realized with a highly reduced computational complexity that increases only linearly with \( N \), i.e., as \( O(C \cdot N \cdot K) \) instead of the original \( O(M^2 T) \) complexity of the one-pass DP algorithm as shown in Fig. 2. Here, \( K \) is the number of pre-quantization segments; \( n \) is the average number of segments from the unit-database having the same label as the labels of a segment in the \( N \)-best lattice and \( C \) is a constant (typically, 10). This enables reaching low distortions with large continuous unit databases unique to unit-selection framework [4] at acceptable complexity.

3. Conclusions

We have proposed a N-best lattice based unit-selection algorithm that has near-optimal R-D performance as the fully optimal one-pass DP algorithm at a highly reduced complexity.

4. References