ZREe report

xrenzh00
ZhiDong Ren

Signal: ESLPod451.WAV
Sample rate:8000Hz
Bit resolution:16bits
Audio channels: mono
Start the project: click project.m

2. Linear prediction
This step will get binary files with features →frame_LPC.mat

3. Code-book creation
This step will get code-book → cb512.txt and gcb512.txt

4. Choose a way to encode the excitation
This step will get binary files with encoded excitation →encoded_excit.mat

5. Test of the code-book
This step will synthesize the speech from quantized LPC and encoded excitation→this speech quality sound normal and the speech length too long lead to be cut.
In the end we synthesized wav files for 1/5 of the recorded read speech.(the result not that bad).

6. Word recognition using DTW
This step : the input and reference matched by DTM and HMM.
Classify: this step finally show us female scores are 4 negative 1 positive, male scores are all positive which means my voice classified into male.

Speech processing

Framing →compute LPC coefficients →compute predicate signal →synthesis predicate signal →synthesis speech signal

LPC Workflow

Speech →output
→Graphic

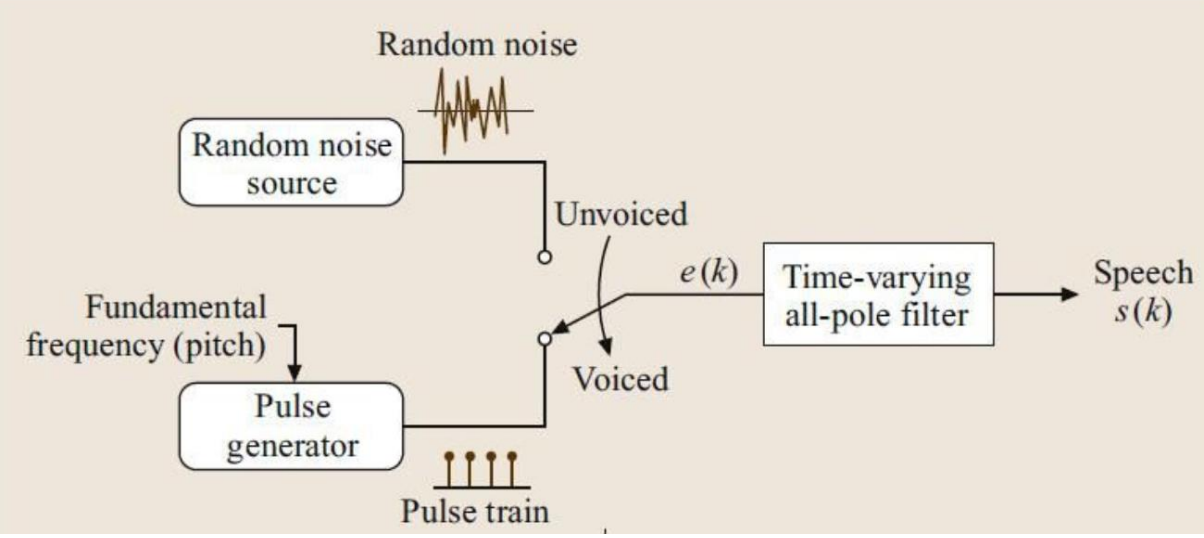
→LPL analysis →LPL coefficients→ →parameter storage
→Pitch analysis→Pitch →parameter storage
→Voiced voiceless decision→parameter storage
Parameter storage →Pitch manipulation &number of coeff →graphic (pitch)
Pitch manipulation &number of coeff →digital filter →output(speech)

Signal var:240000x1
Signal frame:30000x80
Signal mean: 240000x1

Code book

Vector Quantization (VQ) is a lossy data compression method based on block coding rules.
LBG-VQ algorithm solve "multi-dimensional integration"

excitation



Main goal is each frame the excitation is the while Gaussian noise.

Word recognition

Before the mainstream speech recognition system with a large HMM, many scholars used DTW (Dynamic Time Warping) technology for speech recognition. Until now, isolated word recognition for small vocabulary has performed well. The emergence of HMM gradually replaced DTW, but the idea of DTW is that it can be used in many small details of speech recognition.

DTM

DTW can calculate the similarity of two time series, especially for time series with different lengths and different rhythms (an audio sequence where a different person reads the same word). DTW will automatically distort the time series (that is, local zoom on the time axis), so that the morphological fit of the two sequences is consistent, and the maximum possible similarity is obtained.
In time series, the lengths of two time series that need to be compared may not be equal, and in the field of speech recognition, the speech speed of different people is different. Because the voice signal is quite random, even if the same person utters the same tone at different times, it is impossible to have a full length of time. Moreover, the pronunciation speeds of different phonemes in the same word are also different. For example, some people will drag the "A" sound very long or make the "i" sound very short. In these complex cases, the distance (or similarity) between two time series that cannot be effectively obtained using the traditional Euclidean distance.

HMM

Workflow

State set Q with N states
State transition probability matrix
Observation sequence
Observation likelihoods, also called emission probabilities
Initial and end states
forward algorithm--Decoding: Viterbi algorithm--Learning: forward-backward algorithm

Classification

Use of statistical classifiers based on the gaussian distribution