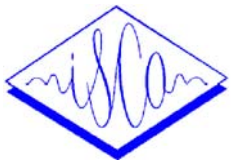

EUROSPEECH '97
5th European Conference on Speech Communication and Technology
Rhodes, Greece, September 22-25, 1997

Referring in Long Term Speech by using Orientation Patterns Obtained from Vector Field of Spectrum Pattern

Kiyoshi Furukawa, Masayuki Nakazawa, Takashi Endo and Ryuichi Oka

Tsukuba Research Center, Real World Computing Partnership
Tsukuba Mitsui building 13F, 1-6-1 Takezono, Tsukuba-shi, 305 Ibaraki, JAPAN
Tel:+81-298-53-1660 FAX:+81-298-53-1740 e-mail:furu@rwcp.or.jp

ABSTRACT

We proposed a new expression of speech feature called orientation patterns which keeps its ability of detection higher in averaging of time domain. Because of this, we achieved to reduce number of frames in reference and input pattern in DP matching algorithm, then the calculation load were reduced.

We constructed long term speech retrieval system by using this new expression. This system has RIFCDP as base matching algorithm which was already proposed. RIFCDP is an algorithm for spotting similar intervals between arbitrary reference pattern and arbitrary input pattern sequence synchronously with input frames.

1.INTRODUCTION

Speech recognition technology is rapidly becoming more sophisticated, and is beginning to be used in the real world. The spotting-based method may be a good basic technique for understanding human spontaneous speech. In usual case, however, plenty of calculation obstruct the system works in real time response.

This paper proposes a new expression of feature of speech which called orientation patterns. The orientation patterns obtained from vector field of spectrum pattern. There is a advantage on using this new expression of speech feature as follows. In continuous DP (CDP) matching using this new feature to detect similar intervals in reference pattern and input pattern, an ability of detection is kept higher in averaging and thinning of time domain. Because of this point, we can make analysis frames longer. Then the number of analysis frames in input and reference of CDP are reduced and a computation load is made lightly using the orientation patterns. We show experimental result of the advantage of the orientation patterns.

We construct speech retrieval system by using the orientation patterns as speech feature. In this system, we used Reference Interval-free Continuous DP (RIFCDP)[2] which is already proposed by us instead of general CDP. The RIFCDP algorithm can calculate and output an adjustment degree between arbitrary intervals in a reference pattern sequence and arbitrary intervals in an input sequence in a spotting-like manner. The speech referring system stores a long term speech as reference, and operator speaks arbitrary word to the system, then the system refers the word in stored reference speech. We can achieve reference interval from long term speech in short time response by using the orientation patterns.

2.RIFCDP

RIFCDP (**Fig.1**) is an algorithm for spotting similar intervals between reference pattern and input pattern sequence synchronously with input frames. The input and reference pattern sequence is therefore assumed to

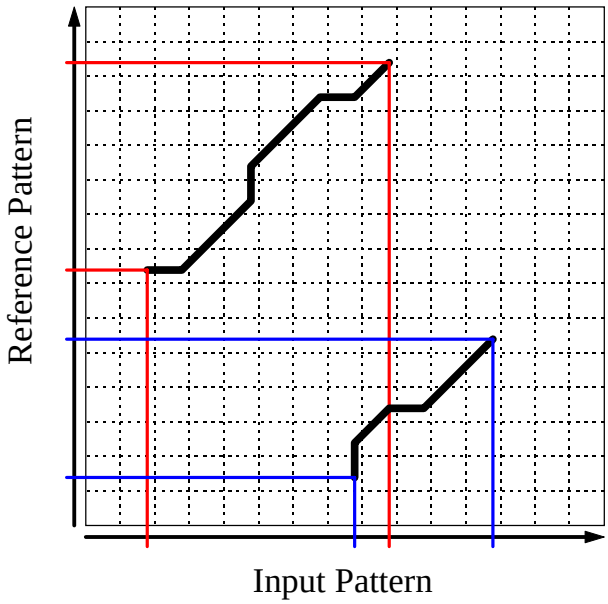

Fig.1: Function of Reference Interval-free Continuous DP (RIFCDP).

non-classified spontaneous speech.

3.ORIENTATION PATTERNS

3.1.Definition

Spectrum vector field (**Fig.2**.b) are obtained from spectrum field between frequency domain and time domain (**Fig.2**.a). The spectrum vector field expresses maximum down slant vector which obtain from to subtract each neighbor point of spectrum field in frequency channel domain and frame number domain.

The orientation patterns (**Fig.2**.c) are obtained from the spectrum vector field. The orientation patterns are considered k dimension scalar vector. The spectrum vector orientation is dropped into a quantumized interval divided to k of 360 degree, the orientation patterns has its absolute value in own category

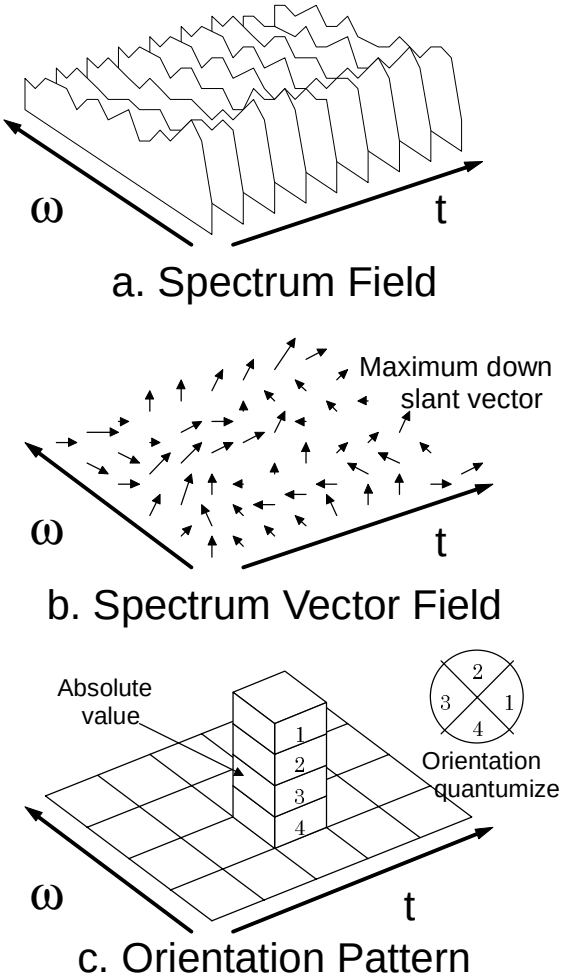

Fig.2: Illustrations of speech feature:
a. Spectrum field
b. Spectrum vector field
c. Orientation pattern.

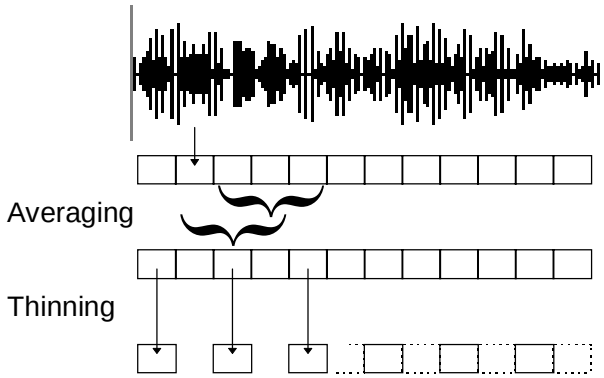

Fig.3: Analysis frame decreasing operation.

correspond to the quantumized interval. Usually the orientation patterns are averaged in time domain in few frames each (**Fig.3**). For example, original analysis frame length in spectrum vector field are 16ms and no overlap each frames, one orientation patterns frame made from one frame. Next, these orientation pattern frames are averaged in 3 frames each, the orientation patterns frame expresses 48ms feature of speech.

3.2.Formulation

This section explain a formulation of orientation pattern. Assuming input speech stream to be *s(t)*, *t* means speech sample sequence number, *q* sec speech stream with *r* Hz sampling rate was expressed *s(t) (0≤t<q·r)*. Assuming analysis frame length to be *L* frames and analysis frame shift to be *m* frames, *n*th analysis frame was made from speech stream from *S(n·m)* to *S(n·m+L-1)*. Spectrum expression of *n*th analysis frame

$$F(n) := \{f(n,i) | 0 \le i < c\} \quad (1)$$

is obtained from FFT with corresponding interval of speech stream from *S(n·s)* to *S(n·s+r-1)*, assuming mel frequency transformation channel to be *c*.

Spectrum vector field *V(n)* *c*-1 dimension vector of vector *v(n,i)* corresponding spectrum *F(n)* was obtained by subtraction operation.

$$V(n) := \{v_t(n,i) | 0 \le i < c-1\} \quad (2)$$

$$v_t(n,i) := (f(n+1,i) - f(n,i),\ f(n,i+1) - f(n,i)) \quad (3)$$

Orientation pattern *O(n)* *c*-1 dimension vector of *k* dimension vector corresponding spectrum vector field *V(n)* was obtained by categorize based on orientation of vector *v(n,i)*, *k* orientation quantumize.

$$O(n) := \{o(n,i) | 0 \le i < c-1\} \quad (4)$$

$$o(n,i) := \{p_{ni}(j) | 0 \le j < k\} \quad (5)$$

$$p_{ni}(j) := \begin{cases} |v(n,i)| & if\ 360/k \le \deg(v(n,i)) < 360(i+1)/k \\ 0 & \text{otherwise} \end{cases}$$

$$\deg((x,y)) := \operatorname{atan}(y/x) \quad (6)$$

3.3.Experiment

The orientation pattern are kept the ability of detection higher in averaging and thinning in order to extend analysis frame length. Generally another expressions of speech feature falls its ability down when extend its analysis frame length.

In experiment, we used two recording speech streams. Both stream contains 24 words. 20 words in these are contained both speech streams and rest 4 words are contained only one side of speech streams. In an ideal case, these 20 corresponding interval pairs of these words contained both speech streams are detected by RIFCDP using one speech stream as reference and the other one as input. Then we got interval pairs in reference stream and input stream each detection. In ideal case, one interval of pair points to word in reference stream and the other interval of pair points to same word in input stream. We evaluated gaps between interval pair of ideal detection and actual detection about each expressions of speech features. We compared between LPC cepstrum, spectrum vector field and two orientation pattern 4 and 8 orientation quantumize.

As preparation we made ideal intervals of each words

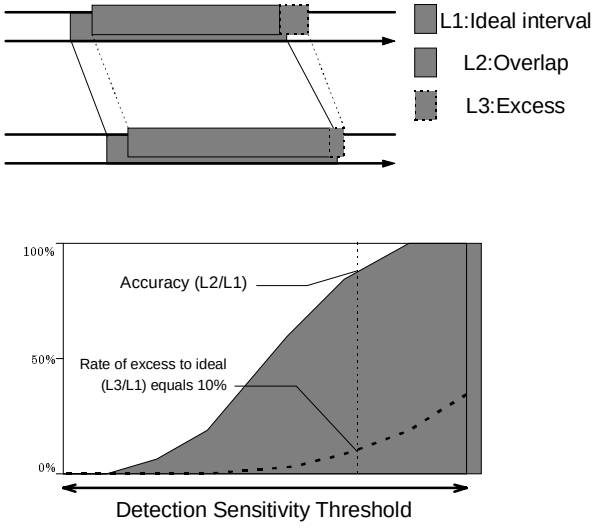

Fig.4: Condition of accuracy evaluation in experiment.

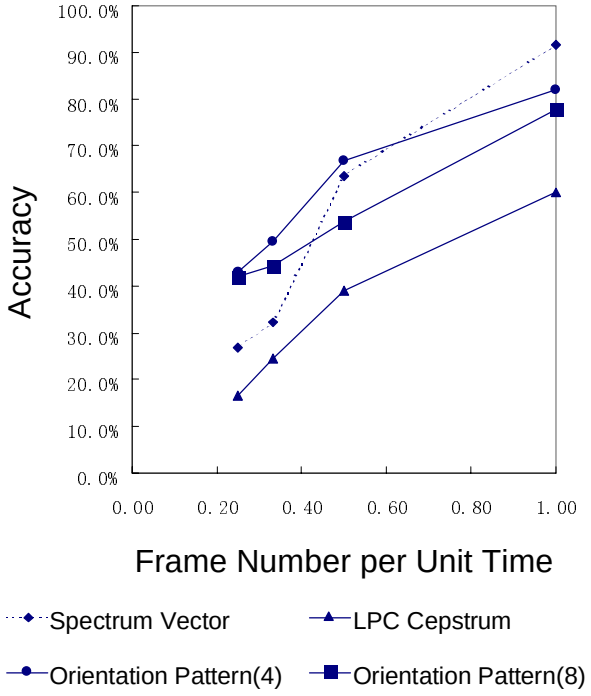

Fig.5: Detection accuracy to frame number per unit time.

about reference stream and input with listening and marking by hand, and we get

L1:totally length of ideal detection intervals.

Detected interval pairs optioned experiment about each features can divide to

L2:overlap to ideal interval,

L3:not overlap to ideal interval.

We set detection sensitivity threshold in order L3/L1*100 to be less than 10%. Under this condition, we evaluated L2/L1*100 as accuracy of interval pair detection. In experiment, analysis frame length was changed from 16ms to 64ms with averaging on time domain. (**Fig.4**)

3.4.Result

These accuracy about each feature was plotted on **Fig.5**. In experiment, the detection accuracy of LPC-cepstrum falls from 60% to 16% when extended its analysis frame length from 16ms to 64ms. Comparing with this, the orientation patterns keeps its ability from 80% to 40% when extended its frame length from 16ms to 64ms.

4.SPEECH RETRIEVAL SYSTEM

We have applied this new feature to speech retrieval

from a database. Since RIFCDP algorithm is frame-synchronous, by regarding a database as a reference pattern, retrieval completes at the time when an utterance input is finished. In a retrieval system, the user should input the retrieval request in spontaneous utterance, and the target intervals that the user wants to extract should be output as soon as the user has uttered enough to identify the target interval.

We assume the words and phrases a user uttered as the request appear and converge in the target interval. Then retrieval of the target interval is considered to be possible to detect the same words and phrases in the database. The RIFCDP algorithm is able to realize these functions in real-time.

We constructed the system which retrievals a voluntary word from recorded long speech like broadcasting, a lecture, a conference referring to the part of voluntary continuous speech. The word which should be retrieved is input by spontaneous speech from a microphone and so on. In order to construct the system, the long speech used as reference is given to reference pattern and the word which an operator input voluntarily is given to input pattern of RIFCDP. A part of common pattern between the reference pattern and input pattern is detected with RIFCDP. The common part is instantly played back in order to make an operator notice that the word has detected as a common part. Moreover, some section that detected part appears more abundantly is chosen and regenerated. The section which includes the voluntary word given by an operator can be played back.

In order to decrease a frame per time, we introduced new feature the orientation patterns. Introduced feature is changed an expression of a feature of a frame many dimensions, it against to averaging and thinning of a frame without falling detective accuracy greatly. As a result, calculative quantity is decreased and detection of the same pair was finished in a frame. We tested the possibility of speech retrieval from a database. A roughly 60 minute speech database was used as the reference pattern. The input was another utterance of the keywords sequence. The interval where the RIFCDP output converges is output at time when the appearance of the output is enough to identify the target.

5.CONCLUSIONS

This paper proposes a new expression of feature of speech which called orientation patterns. An ability of detection of this new feature is kept higher in averaging of time domain. Then the number of analysis frames in CDP are reduced and a computation load is made lightly. We shown experimental result of the advantage of the orientation patterns.

We construct speech retrieval system by using the orientation patterns as speech feature. The speech retrieval system stores a long term speech as reference, and operator speaks arbitrary word to the system, then the system refers the word in stored reference speech. We can achieve reference interval from long term speech in short time response by using the orientation patterns.

- propose new expression of speech feature which has robustness on averaging in time domain.
- show experimental result as advantage of the robustness.
- construct long term speech referring system to reduce analysis frames in reference using by orientation patterns as feature.

Reference

[1] K.Furukawa, M.Nakazawa, J.Kiyama, Y.Itoh and R.Oka, “Speech retrieval and speech summary,” Proc. of Real World Computing Symposium ’97, Japan, 1997.

[2] Y.Itoh, J.Kiyama, H.Kojima, S.Seki and R.Oka, “A proposal for a new algorithm of reference interval-free continuous DP for real-time Speech or text retrieval,” Proc. of ICSLP'96, 1996.

[3] Y.Itoh, J.Kiyama and R.Oka, “Speech understanding and speech retrieval for TV news by using connected word spotting,” Proc. of Eurospeech’95,1995.

[4] R.Oka and H.Matsumura, “Speaker-independent word speech recognition using the blurred orientation pattern obtained from the vector field of spectrum,” Proc. of 9th Int. Joint Conf. on Pattern Recognition, 1988.

[5] R.Oka, “Phonemic recognition of each frame with vector field feature using continuous dynamic programing”, IEEE, Proc. of ICASSP’86, 1986.