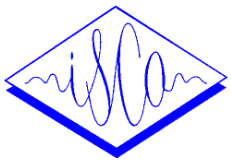


First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

LOMBARD SPEECH RECOGNITION BY FORMANT-FREQUENCY-SHIFTED LPC CEPSTRUM

Yumi TAKIZAWA Masahiro HAMADA

Central Research Laboratories,
Matsushita Electric Industrial Co., Ltd.
3-15, Yagumo-Nakamachi, Moriguchi, Osaka 570 JAPAN

ABSTRACT

To perform speech recognition in a noisy environment, it is important to minimize the influences of both additive noise and the Lombard effect. Data analysis clearly showed a trend of formant frequency shift in some specific frequency bands due to the Lombard effect. This paper reports a method of modifying the LPC cepstrum to compensate for the formant frequency shift, which had usually been left uncorrected. The compensated value can be obtained by multiplication of the formant frequency shift value and the partial differential of the LPC cepstrum with respect to the formant-frequency. To validate the effect of the compensation method proposed herein, the following items were studied.

(1) Comparison to the usual method (Compensation of spectral tilt).
(2) Comparison to the weighted cepstrum.

I. INTRODUCTION

It has been already confirmed that the Lombard effect largely degrades the performance of the speech recognition systems [1],[2],[3]. As countermeasures to minimize the effect, several ideas were proposed for instance, a method to update the reference speech [2],[4],and a training method to properly educate the speakers to avoid Lombard speech in noisy environments[5].

On the other hand, as the result of acoustical analysis of Lombard-speech data, it became clear there was a specific trend only seen in Lombard speech in the speech amplitude, pitch, duration, spectral tilt, formant-frequency, etc. [6],[7],[8],[9]. Taking above analysis results into account, methods proposed in recent years include cancelling change of spectral tilt caused by the Lombard effect by utilizing least square error curve[10] and slope-dependent weighting [3], and use of a LPF to suppress high band frequencies, which largely vary due to the Lombard-effect[10]. However, no compensation method for the specific trend in formant frequency has been proposed as yet.

This paper, after making clear the degree of influence from the Lombard effect and additive noise through recognition experiment, verifies that formant frequency does shift by the Lombard effect, and proposes a compensation method of the above frequency shift in the LPC cepstrum which is a recognition parameter vector. And, through speaker dependent word recognition experiments, the effect of the proposed compensation is verified.

Table 1 Feature of recognizer and database

Recognizer
10 KHz sampling
14 order LPC cepstrum
Frame length 25.6 msec
Frame shift 12.8 msec
Euclidean distance
Database (By courtesy of Speech Tech. Lab.)
11 degits, 14 speakers(9 females and 5 maales)
Database were recorded two repetitions for both noiseless and noisy environments.
"Noisy envionments" means that the speakers were listening to white-Gaussian noise at 85 dBSPL through calibrated headphones.
Additive noise
pink noise (-6dB/oct), SNR 0dB

II. THE LOMBARD EFFECT ON RECOGNITION

Simulating speech inputs in both noiseless and noisy environmental conditions, recognition experiments were performed by changing reference speech conditions. Also, recognition rate changes were studied using cepstral coefficients weighted by RPS [11] which is robust for additive noise. Table 1 shows features of the recognition system and data base and additive noise.

Results of the experiment are shown in Tables 2 and 3. When an utterance was spoken under the "Lombard + Noise" condition, the speech recognition rate was higher in using "Lombard speech with no noise" as reference than "non-Lombard speech with additive noise". In this experiment, the influence from the Lombard effect is greater than that of additive noise. (Compare recognition rate of [1] and [2] in Table 2.)

By using the weighted cepstrum, the above mentioned trend appears more clearly. The weighted cepstrum is effective for additive noise, but not so much for the Lombard effect. (Compare recognition rate of [3] and [4] in Table 3.)

Table 2 Recognition for Lombard data

Input/Ref.	Non-Lom.	Lom.	Non-Lom. +Noise	Lom. +Noise
Non-Lom.	96.4% 94.5%	79.1% 78.2%	76.4% 92.1%	58.1% 78.1%
Lom. +Noise	27.3% 39.0%	76.4% 83.6% [1]	74.5% 78.1% [2]	96.4% 96.4%

Upper row : Rate for Male
Lower row : Rate for Female
Lom. : Lombard

Table 3 Recognition used RPS lifter for Lombard data

Input/Ref.	Non-Lom.	Lom.	Non-Lom. +Noise	Lom. +Noise
Non-lom.	98.1% 98.1%	78.2% 90.9%	92.7% 96.4%	78.1% 88.1%
Lom. +Noise	50.0% 78.1%	90.9% 96.4% [3]	72.7% 78.1% [4]	98.1% 98.1%

III. ACOUSTIC CHARACTERISTICS OF LOMBARD SPEECH

Acoustic analysis of steady vowel portions in Lombard speech data was carried out. Fig.1 and Fig.2 show analysis results regarding formant frequency and LPC cepstrum coefficients. The values plotted in drawing indicate average values in the steady vowel portions.

A. Formant frequency

In the case of Lombard speech, without correlating to phoneme, both the first and second formants shift to higher frequencies when they are below about 1.5kHz. The average value of the shift is about 120Hz. It was comfirmed the average value of the shift is larger than the standard deviation of shift in most of phonemes. The formants above about 1.5kHz are inclined to shift to a lower frequency. The higher the formant frequency, the less the degree of resultant shift.

B. LPC cepstral coefficient

Coefficient values tend to be smaller in Lombard speech. This trend is more marked for lower cepstral order.

IV. FORMANT-FREQUENCY-SHIFTED LPC CEPSTRUM

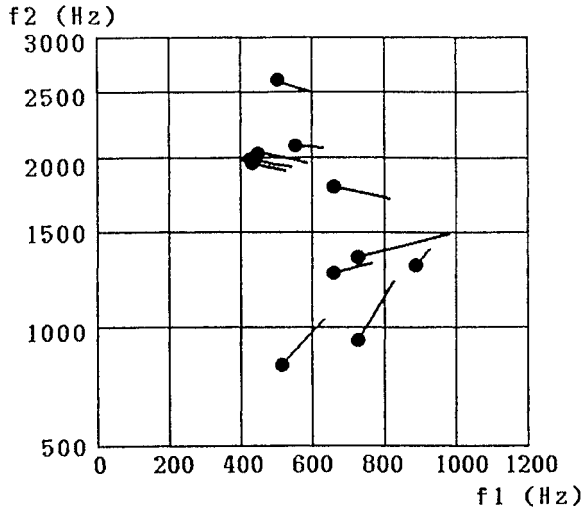
The authors propose a method of effecting formant frequency compensation by modifying the LPC cepstrum coefficients. This can be done by examining the relationship between the formant frequency shifts and the LPC cepstrum shifts.

A. Compensation method

We'd like to propose a compensation formula(1) to compensate cepstral coefficient values in Lombard speech to coefficient value in non-Lombard speech. Compensation value γ_n in formula (1) is equal to the change in that coefficient when the formant frequency is shifted. And, this is proportional to the partial derivative of the LPC cepstral coefficient with respect to formant frequency and shift value Δf in formant frequency. Therefore, the compensation value can be defined by formula (2).

$$\tilde{C}_n = C_n + \gamma_n \qquad (1)$$

where :
$\tilde{C}_n$: LPC cepstral coefficient after compensation
C_n : n-th LPC cepstral coef. of Lombard speech
γ_n : Compensation value of C_n



• : Non-Lombard
—— : Direction of the shift from non-Lombard to Lombard

Fig.1 Formant frequency shifts from non-Lombard to Lombard

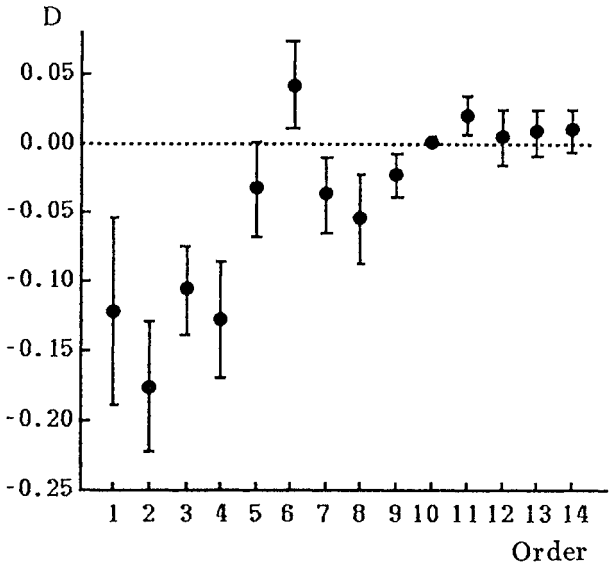


Fig.2 Difference averages and variance between cepstrum coefficient values in Lombard speech and non-Lombard speech, ($D = C_i^n - C_i^n$)

$$\gamma_n \equiv \sum_{i=1}^{M/2} (\Delta f \ \partial \frac{C_n}{\partial f_i}) \qquad (2)$$

where :

Δf : Difference in formant frequency between Lombard speech and non-Lombard speech

f_i : i-th formant frequency of Lombard speech

On the other hand, each cepstral coefficient can be expressed by the formula (3) using all formant frequencies. That is to say, each cepstral coefficient can be expressed by the summation (according to the number of pole pairs) of the product between 1) an exponential function which variable is band width of i-th formant frequency and 2) a cosine function whose variable is the i-th formant frequency. Therefore, the contents of the partial differential equation of formula(2) becomes formula(4).

$$C_n = \frac{2}{n} \sum_{i=1}^{M/2} (exp(\frac{-n\pi b_i}{K}) \ cos(\frac{2\pi f_i n}{K})) \qquad (3)$$

where :

b_i : Band width of i-th formant

K : Sampling frequency

$$\partial \frac{C_n}{\partial f_i} = - \frac{4\pi}{K} \ exp(\frac{-n\pi b_i}{K}) \ sin(\frac{2\pi f_i n}{K}) \qquad (4)$$

Practically, average value of this data Δf=120Hz, all b_i=150Hz, and we take 10,000 as the sampling frequency. Only for the range of 300Hz to 1,500Hz where formant frequency differences were recognized, apply the above compensation formula. The compensation value can be expressed by the following formula (5).

$$\gamma_n \equiv -0.15 \ exp(-0.047n) \sum_i (sin(\frac{2\pi f_i n}{10000})) \qquad (5)$$

where : $300 < f_i < 1{,}500$

B. Evaluation of compensation formula

To validate the effect of the above mentioned compensation formula, the authors intentionally shifted formant frequency, and checked the equality between 1) the difference of each LPC cepstral coefficient value before and after shifting formant frequency and (2) the compensation value calculated by the compensation formula.

The solid points in Fig.3 show the differences (measured value) of LPC cepstral coefficient values before and after formant frequency shift, when only the second formant (1325Hz) of vowel/a/ is shifted Δf=120Hz. The dotted line in Fig.3 shows the compensated value (theoretical value) calculated by inserting the value (1325+60=1385Hz) into formula (5) obtained by adding 1) the above mentioned formant frequency (1325Hz) and 2) the shift value f/2(120/2=60Hz)

The difference of measured coefficient values before and after formant shifting was a sine-wave characteristic. This sine-wave cycle which correlates to formant frequency corresponds to the cycle of theoretical compensation value. The first formant of vowel/i/ was also validated in the same manner.

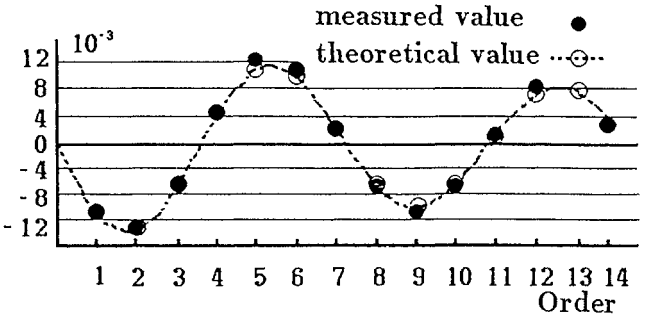


Fig.3 Relation between LPC cepstral coefficient value shifts according to /a/ 2nd formant shift (measured values) and theoretically compensated values

V. EXPERIMENTS AND RESULTS

A. Comparison to the usual method

(a) Method

The authors applied the above compensation to Lombard speech data and validated the compensation effect. In the same manner as [10], after applying a 2.5 kHz LPF on the data, the experiment was carried out by comparing 1) the usual method: varying the amount of pre-emphasis, and 2) the proposed method according to formula (5). In addition, the effect when combining the usual method and the proposed methods ("combined method") was also validated. The above mentioned pre-emphasis control method is to set the pre-emphasis amount to 0.6375 only for vowel portions of Lombard speech and 0.9375 for other portions. The recognition rate appeared to be higher for our data, when fixing the pre-emphasis at 0.6375, than adapting the pre-emphasis. Therefore, used the fixed amount of pre-emphasis.

(b) Result

In this experiment, Lombard speech data was used as input and non Lombard speech was used as reference. Table 4 show recognition rates without noise added to the input voice (experiment 1) and with noise added (experiment 2) respectively. The recognition rate when non Lombard speech was used as input voice was described as the target recognition rate.

In experiments 1 and 2, compensation effects of the usual method and the proposed method were validated ; however, when average among all speakers, the usual method is more effective than the proposed one. Thus, we suppose that influences from both high-order formant distortion and mid-band power increase are greater than the influence affected the recognition by the formant frequency shift.

However, looking at the recognition rates for individual speakers, the proposed method consistently improved the recognition rate, while the usual method degraded some. The authors confirmed that the cause was lack of voice information due to LPF.

Compensation effect increases when the usual and the proposed methods are combines, as opposed to separate. This is because the elements of compensation differ in the usual and proposed methods. The combined method of the usual and proposed methods is superior to the separate method because both help each other overcome the insufficiency.

Table 4 Recognition rates according to various correction methods

Applied process	Lombard without noise	Lombard with noise
Without process	69.9 %	32.3 %
Usual method	79.0 %	53.9 %
Measured value used	73.0 %	40.3 %
Proposed method	75.6 %	46.0 %
Combined method	83.0 %	60.2 %
Target	93.0 %	67.2 %

Useal method	LPF plus spectral tilt compensation.
Measured value used	Compensated by the average value (calculated in each order)of cepstral coefficient differences according to Lombard speech.
Proposed method	Compensation of formant frequency.
Combined method	Usual plus proposed method.
Target	When there's no difference of speech between the reference and input.

B. Combined use with weighted-cepstrum coefficient

Correction effect was studied when combined with cepstral coefficients weighted by RPS. Table 5-1 and 5-2 show the result of this study. Recognition rate increases when compensation carried out independently from additive noise and RPS weighting.

In addition, the proposed compensation method provides greater correction effect when combined with RPS weighted cepstral coefficients. The high-order cepstrum weighting emphasizes detailed information such as formant position, rather than figurative information such as spectrum tilt. The proposed method improves the reliability of the above mentioned detailed information by compensating formant frequency, therefore the recognition rate is expected to increase more by high-order cepstrum weighting.

Table 5-1 Recognition rate for Lombard without noise

	Not weighted	Weighted
Not compensated	66.9 %	72.7 %
Compensated	75.6 %	81.8 %

Table 5-2 Recognition rate for Lombard with noise

	Not weighted	Weighted
Not compensated	32.3 %	64.2 %
Compensated	46.0 %	68.6 %

VI. CONCLUSIONS

(1) In this experiment, the influence of the Lombard effect on the recognition rate was greater than that of additive noise; also, weighted cepstrum was not effective for the compensation of speech transformation.

(2) It was observed that if formant frequency was lower than about 1.5 kHz, the formant frequency was shifted up by the Lombard effect; if higher than about 1.5 kHz, it was shifted down. Also, by the Lombard effect, the LPC cepstrum coefficient values tends to decrease; this tendency becomes more significant in the lower order of coefficients.

(3) When compensating the above-mentioned formant frequency shift via the LPC cepstrum, it was found that the compensation was possible by using the value of partial-differential of coefficient with respect to formant frequency. We did confirm the compensation effect.

(4) This proposed compensation method differs from the usual one in what it compensates for. The combined use of both the proposed and usual methods enhances the effect.

(5) Also, the proposed correction method enhances the compensation effect by combined use of RPS weighted cepstral coefficients.

REFERENCES

[1] Rajasekaran,P.K. "Recognition of speech under Stress and in Noise" Proc. of ICASSP86, pp 733-736, 1986

[2] Roe,D. "Adaptation of a Speech Recognizer to the Lombard Effect in High Noise Conditions" IEICE Technical Report SP86-66 pp 41-48, 1986

[3] Stanton,B.J. "Robust Recognition of Loud and Lombard Speech in the Fighter Cockpit Environment" Proc. og ICASSP89, PP 675-678, 1989

[4] Baker,J.M. "Optimal and suboptimaltraining Strategies for Automatic Speech Recognition in Noise, and the Effects of Adaptation of Performance" Proc. of ICASSP86, pp 745-748, 1986

[5] Herbert,L.P. "Inhibiting the Lombard Effect" J.Acoust.Soc.Am. vol.85 No.2 pp 894-900, Feb. 1989

[6] Stanton,B.J. "Acoustic-Phonetic Analysis of Loud and Lombard Speech in Simulated Cockpit Conditions" Proc. of ICASSP88, pp 331-334, 1988

[7] Pisoni,D.B. "Some Acoustic-Phonetic Correlates of Speech Produced in Noise" Proc. of ICASSP86, pp 1581-84, 1986

[8] Summers,W.V. "Effects of Noise on Speech Production: Acoustic and perceptual analyses" J.Acoust.Soc.Am. Vol.84 No.3 pp 917-928, Sep. 1988

[9] Hanson,J.H.L. "Evaluation of Acoustic Correlates of Speech Under Stress for Robust Speech Recognition" Proc. Annu Northest Bioeng Conf Vol.15 pp 31-32, 1989

[10] Hattori H. "A study for Speech Recognition Under Noisy Environment" IEICE Technical Report SP88-11 pp 1-6, 1988

[11] Hanson,B.A. "Spectral Slope Based Distortion Measures for All-Pole Models of Speech" Proc. of ICASSP86, pp 757-760, 1986

[12] Itahashi S. "On Properties of Speech Cepstra" EIC (D) J71-D,9, pp 1839-42, 1988