

# 바람잡음을 고려한 자동차에서의 음성인식 성능 향상

이 기 훈, 이 철 희, 김 중 교

전북대학교 전자 공학과

## Improvement of Speech Recognition Performance in Running Car by Considering Wind Noise

Ki-Hoon Lee, Chul-Hee Lee, Chong-Kyo Kim

Department of Electronic Engineering, Chonbuk National University

khlee@ssplab.chonbuk.ac.kr

### Abstract

This paper describes an efficient method for improving the noise-robustness in speech recognition in a running car by considering wind noise. In driving car, mainly three kind of noises engine noise, tire noise and wind noise, are severely affect recognition performance. Especially wind noise is an important factor in driving car with window opened. We analyzed wind noise in various driving conditions that are 60, 80, 100 km/h with window fully opened, window half opened. We clarified that the recognition rate is significantly degenerated when the wind noise components in the frequency range above 200 Hz are large. We developed a preprocessing method to improve the noise robustness despite of wind noise. We adaptively changed the cutoff frequency of the front-end high-pass filter from 100 through 200 Hz according to the level of the wind noise components. By this method, the recognition rate is considerably improved for all kind of driving conditions

진 잡음, 타이어 잡음 그리고 바람잡음(wind noise)에 의한 인식성능의 저하를 해결하는 것이다. 지금까지 많은 연구를 통해 주행하는 자동차에서 음성인식 성능은 현저한 향상결과를 보여 왔다[1-3]. 그러나, 창문을 열고 주행하는 자동차에서의 음성 인식에 관한 연구는 거의 이루어지지 않고 있다.

본 논문은 창문을 열고 주행하는 자동차에서 음성인식 성능 저하에 가장 중요한 요인 중의 하나인 바람잡음의 특성을 분석하여 바람잡음(wind noise) 환경에서도 효율적인 음성인식을 위한 방안을 제안한다. 창문을 열고 주행하는 자동차에서 바람잡음(wind noise)은 일반적으로 200 Hz 이하의 주파수 영역에 분포한다. 본 논문에서는 주행 속도에 따른 서로 다른 신호 대 잡음비(signal to noise ratio)를 이용하여 차단주파수(cut-off frequency)를 다르게 설정하는 AHPF(adaptive high pass filter)와 음성인식 향상을 위해 NSS(noise spectral subtraction)를 적용하였다.

본 논문은 2장에서 자동차 주행속도에 따른 wind noise 특성을 분석하고, 3장에서는 HPF(high pass filtering)와 NSS(noise spectral subtraction)에 대해서 논하며, 4장에서는 적응 잡음제거를 하기 위한 AHPF(adaptive high pass filtering)에 대해서 설명한다. 5장에서는 실험방법 및 결과에 대해서 논하고, 6장에서 결론을 맺는다.

### 1. 서 론

최근 음성인식과 관련하여 자동차 네비게이션 시스템과 같은 운전자의 안정성과 편의성을 향상시키고자 하는 연구가 활발히 진행되고 있다. 자동차에서 음성인식성능을 향상시키기 위해서 가장 중요한 문제는 엔

### 2. 음성과 잡음 데이터

20명의 화자가 100개의 단어를 발성하였고, 16 kHz 로 샘플링 되었으며, 16 bit 로 양자화 된 데이터를 이

용하였다. 잡음데이터는 르노 삼성 SM5(2000cc)을 이용하여 창문을 완전 개방한 상태(fully opened)와 반절 개방한 상태(half opened)에서 60, 80, 100 (km/h)의 속도로 주행하는 자동차에서 핸즈프리 마이크를 통해서 수집하였다.

그림 1은 창문을 완전 개방하고 여러 가지 속도(60, 80, 100 (km/h))로 주행하는 자동차에서 수집한 wind noise의 주파수 영역에서의 power spectrum을 나타낸다. 일반적으로 음성에서 중요한 정보를 갖는 신호의 주파수는 200 Hz 미만에 분포가 거의 없는 반면[2]에 wind noise의 주파수는 60 km/h에서 100 Hz 미만, 80 km/h에서 150 Hz 미만, 그리고 100 km/h로 주행중인 자동차에서는 200 Hz 미만에 분포한다는 것을 알 수 있다

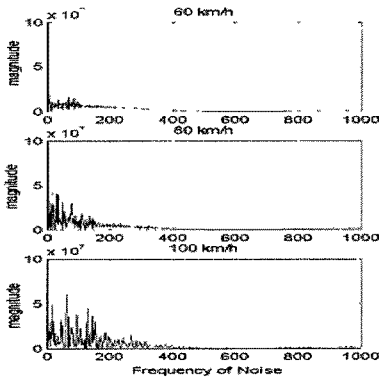


그림 1 Wind noise power spectrum

그림 2에서는 실제 주행하는 자동차에서 wind noise  $n(k)$ 를 수집하여 clean speech  $s(k)$ 에 첨가한 파형을 보여준다. 여기서 noisy speech  $y(k)$ 는 식 (1)과 같다.

$$x(k) = s(k) + n(k) \quad (1)$$

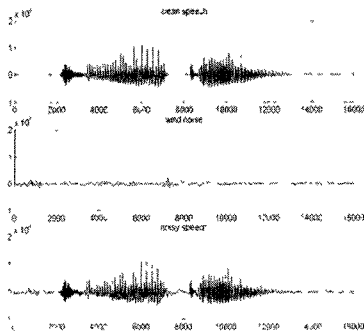


그림 2 wind noise가 첨가된 음성 파형

### 3. 전처리 기법

#### 3.1. HPF(high pass filtering)

자동차에서 음성인식의 성능을 저하시키는 요인으로 wind noise, 엔진잡음과 노면상태에 따른 tire noise가 있다. 이러한 잡음들은 일반적으로 500 Hz 이하의 주파수 영역에 대부분의 성분들이 분포되어 있다[2]. 따라서 잡음의 레벨에 맞는 적당한 차단주파수를 이용하여 HPF(high pass filtering)을 하는 방법을 많이 이용해 왔다.

본 논문에서는 wind noise의 주파수 성분들이 100 ~ 200 (Hz) 사이에 분포하므로 전처리 과정에서 실험에 따라 차단주파수를 설정하여 high pass filtering을 수행하였다 사용된 필터는 식 (2)의 주파수 응답을 갖는 9차의 butterworth filter를 사용하였다. 여기서,  $a$ ,  $b$ 는 길이가  $n+1$ 인 필터계수이다.

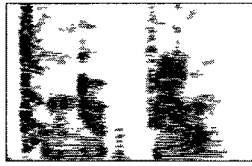
$$H(z) = \frac{B(z)}{A(z)} = \frac{b(1) + b(2)z^{-1} + \dots + b(n-1)z^{-(n-1)} + b(n)z^{-n}}{1 + a(2)z^{-1} + \dots + a(n+1)z^{-n}} \quad (2)$$

#### 3.2. NSS(noise spectral subtraction)

전처리 과정에서 잡음을 제거하는 기법 중에서 간단하고 구현이 용이하여 널리 쓰이는 NSS는 noisy speech의 power spectrum  $Y^2(m, f_k)$ 에서 잡음의 power spectrum  $N^2(m, f_k)$ 을 제거함으로써 원음성(clean speech)의 power spectrum  $S^2(m, f_k)$ 를 구하는 것으로서 식 (3)과 같이 표현할 수 있다. 여기에서  $m$ 은 power spectrum상에서 프레임의 값이고,  $f_k$ 는  $k$ 번째 spectral component를 나타낸다.

$$S^2(m, f_k) = Y^2(m, f_k) - N^2(m, f_k) \quad (3)$$

그림 3은 100 km/h의 wind noise 데이터를 첨가한 noisy speech에 HPF와 NSS를 적용한 spectrogram을 나타냈다.



(a)



(b)

그림 3 Noisy speech에 대한 spectrogram

(a) NSS 적용 전 (b) NSS 적용 후

Noise power spectrum을 모를 경우에는 식 (4)를 이용하여 추정된 noise power spectrum  $\hat{N}(m, f_k)$ 을 이용하여 noisy speech에서 잡음을 제거한다.

$$H(m, f_k) = 1 - \left( \frac{\hat{N}(m, f_k)}{Y^2(m, f_k)} \right)^{1/2} \quad (4)$$

## 4. 적응 잡음제거

### 4.1. AHPF(adaptive high pass filtering)

주행중인 자동차에서 속도에 따른 잡음의 영향을 보상해주기 위해서, 전처리 과정에서 wind noise 성분의 주파수 분포에 따른 100 ~ 200 (Hz)의 차단주파수를 갖는 고역통과 필터를 이용하였다.

차단주파수는 wind noise와 음성 데이터들의 에너지 영역에서의 SNR을 이용하여 선정하였으며, SNR은 식 (5)를 통해 오차(error) 신호의 에너지와 음성(clean speech) 신호의 에너지를 이용하여 구할 수 있다. 그림 4는 각각 창문을 완전 개방한 상태(f60, f80, f100)와 절반 개방한 상태(h60, h80, h100)의 10개의 데이터에 대해서 SNR의 변화를 보여준다.

$$SNR = 10 \log_{10} \frac{E_s^2}{E_e^2} = 10 \log_{10} \frac{\sum s^2(k)}{\sum [s(k) - \hat{s}(k)]^2} \quad (5)$$

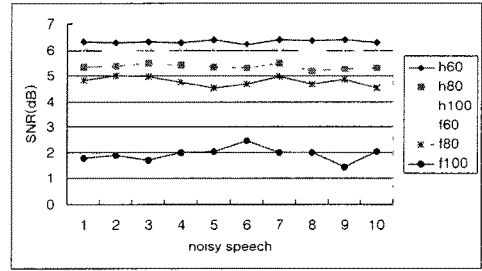


그림 4 Noisy speech에 대한 SNR

본 논문에서는 창문을 열고 주행하는 자동차 내에서 음성인식 성능향상을 위해 잡음의 조건에 대한 SNR을 이용하여 100, 150, 200 Hz의 차단주파수를 이용한 AHPF(adaptive high pass filtering)을 제안하였다. 차단주파수는 식 (6)에 의해서 선정하였으며, 각 주행조건에 따라 SNR과 차단주파수 ( $f_{cutoff}$ )를 나타내면 표 1과 같다.

$$f_{cutoff} = \begin{cases} 100 \text{ Hz} & 5.6 \text{ dB} < \text{SNR} \\ 150 \text{ Hz} & 4 \text{ dB} < \text{SNR} \leq 5.6 \text{ dB} \\ 200 \text{ Hz} & 4 \text{ dB} \geq \text{SNR} \end{cases} \quad (6)$$

표 1. 자동차 주행 조건에 대한 SNR 과  $f_{cutoff}$

주행조건	f60	f80	f100	h60	h80	h100
SNR(dB)	6.32	5.35	3.41	6.01	4.77	2.03
$f_{cutoff}$ (Hz)	100	150	200	100	150	200

## 5. 실험 방법 및 결과

HTK 3.2.1을 이용하여 화자독립 고립단어 인식 실험을 수행하였다. 20명의 화자가 각각 100개의 단어를 발성하였으며, 이중 15명을 훈련용(training)으로 사용하였고, 5명을 테스트(test)용으로 총 2000개의 데이터를 사용하였다.

실험①에서는 60, 80, 100 (km/h)의 wind noise를 clean speech에 첨가하여 인식실험을 하였으며 그 결과를 표 2, 3의 실험①에 나타냈다.

실험②에서는 전처리 과정에서 일괄적으로 150 Hz의 차단주파수를 모든 noisy speech에 대해서 HPF(high pass filtering)를 거치고 NSS(noise spectral subtraction)를 수행한 인식실험을 하였고, 그 결과를 표 2, 3의 실험②에 나타냈다.

실험③에서는 식 (6)에 따라 noisy speech 각각의 SNR값에 따른 100, 150, 200 (Hz)의 적응 차단주파수를 선정하여 AHPF를 거친 후 NSS를 적용하였다. 표 2, 3과 그림 5에서 볼 수 있듯이 주행하는 자동차에서 주행속도에 따라서 변하는 wind noise의 특성을 고려한 실험③의 AHPF(adaptive high pass filtering)방법이 실험②보다 향상된 결과를 보였다.

표 2. 창문 완전 개방시 인식실험 결과

		창문 완전 개방상태		
		60 km/h	80 km/h	100 km/h
실험①	Noisy Speech	70.89 %	68.44 %	68.11 %
실험②	HPF + NSS	76.44 %	77.49 %	77.67 %
실험③	AHPF + NSS	83.67 %	77.51 %	83.33 %

표 3. 창문을 절반 개방시 인식실험 결과

		창문 반절 개방상태		
		60 km/h	80 km/h	100 km/h
실험①	Noisy Speech	73.89 %	70.44 %	68.56 %
실험②	HPF + NSS	77.50 %	78.95 %	78.44 %
실험③	AHPF + NSS	84.40 %	78.99 %	82.58 %

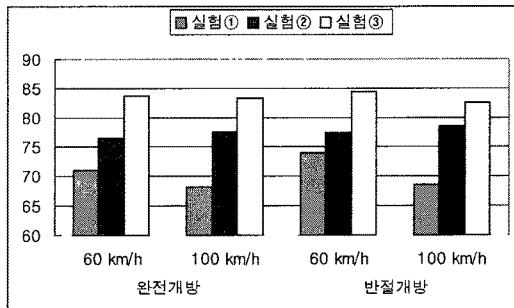


그림 5. 인식실험 결과

## 6. 결 론

본 논문에서는 창문을 열고 주행중인 자동차에서 음성인식성능을 향상시키기 위해 wind noise의 특성을 고려한 적응 고역통과필터를 이용하였다. 또한 효과적인 간단한 잡음제거 기법으로 알려진 NSS(noise spectral subtraction)과정을 거쳐 전처리 단계에서 음성인식의 성능을 저하시키는 잡음 성분을 제거해주

었다. 인식실험 결과 음성신호에 wind noise를 첨가하여 일괄적인 차단주파수(150 Hz)를 적용한 후 NSS(noise spectral subtraction)를 수행한 전처리 기법보다 주행하는 자동차의 wind noise의 특성을 분석하여 SNR에 따라 차단주파수를 설정한 AHPF(adaptive high pass filtering)과정을 적용한 인식실험 결과가 모든 환경에서 개선된 결과를 보여주었다.

향후 계획으로는 자동차 운전시 발화자의 입의 위치와 창문을 통한 공기의 흐름의 특성을 고려한 공기역학적인 관점에서의 연구와, 효과적인 음성인식을 위해서 non-steady noise에 대한 보다 구체적인 연구가 필요하다.

## 참 고 문 헌

- [1] Lecomte I., Lever M., Boudy J, and Tassy A., "Car Noise Processing for Speech Input," IEEE Proc. ICASSP, vol. 1, pp. 512-515, 1989.
- [2] Mokbel C. E. and Chollet G. F. A, "Automatic Word Recognition in Cars," IEEE Trans. Speech and Audio Process., vol. 3, pp. 346-356, 1992.
- [3] Shozakai M., Nakamura S. and Shikano K., "Robust Speech Recognition in Car Environments," IEEE Proc. ICASSP, vol. 1, pp. 269-272, 1998.
- [4] Boll, S. F, "Suppression of Acoustic Noise in Speech using Spectral Subtraction," IEEE Trans. Acoust. Speech signal Process., vol. 28, pp. 113-120, 1979
- [5] Stefanie Aalborg, Christophe Beaugeant, Sorel Stan, Tim Fingscheidt, Radu Balan, and Justinian Rosca, "Single- and two-channel Noise Reduction for Robust Speech Recognition in Car," Siemens Corporate research, Multimedia and Video technology, 2002.
- [6] Steve Young, Gunnar Evermann, "HTK Book" version 3.2.1, Cambridge University Engineering Department, 2003.