잡음 환경에 강인한 원거리 음향 정보 검출 기술 연구

유 인 철, 육 동 석
고려대학교 컴퓨터학과

Noise robust distant sound recognition

Inchul Yoo, Dongsuk Yook
Department of Computer and Communication Engineering, Korea University
E-mail : icyoo@voice.korea.ac.kr, yook@voice.korea.ac.kr

Abstract

This paper reviews the issues in implementing sound recognizers in real environments. First is the signal corruption caused by background noises and reverberation. Second is the open-set problem which is the problem of rejecting out-of-vocabulary words and noises. These two issues must be solved for noise robust recognizers.

I. 서론

청각 장애인들은 일상 생활에서의 전화벨 소리나 초인종 소리 등 중요한 음향 정보에 대한 접근이 어렵다. 청각 장애인을 위한 보청견은 훈련과 유지에 많은 비용과 시간이 소요되며, 사용 가능한 환경에 제약이 따르게 된다. 따라서 음성 인식 기술을 응용하여 컴퓨터로 이들 보청견의 기능을 대체할 수 있는 음향 정보 인식 시스템을 구축할 필요가 있다.

실제 환경에서는 통제된 실험실 환경과는 달리 많은 잡음이 존재하며 실내의 반향으로 인해 소리가 왜곡이 발생한다. 또한 구현하고자 하는 시스템은 주변의 소리를 항시 받아들이고 있어야 하므로 등록되지 않은 음향에 대해서 잘못 반응하지 않아야 한다. 본 논문에서는 이러한 실내 음향 왜곡 문제와 미등록 음향의 거절에 관하여 알아보도록 한다.

II. 실내의 음향 왜곡문제

실생활에서는 여러 주변 잡음과 실내의 반향으로 인해 음향이 왜곡된다. 이렇게 왜곡된 음향신호는 인식기를 학습할 때 사용하였던 데이터와는 다른 모습을 가지게 되어, 성능에 큰 저하가 발생한다. 이러한 문제를 해결하기 위해서는 크게 두 가지 방법이 있다. 첫 번째는 왜곡된 음향과 인식기의 학습 데이터 상의 차이를 보완하는 방법으로 음향의 왜곡을 상쇄하거나 인식기를 환경 왜곡에 적응시키는 등의 기법을 사용하게 된다. 두 번째 방법은 이러한 왜곡의 영향을 최대한 적게 받는 특징 추출 혹은 유사도 측정 알고리즘의 개발이다.

음향 왜곡 문제를 해결하기 위한 알고리즘 중 가장 널리 알려진 방법은 스펙트럼 차감법 (spectral subtraction) [1] 이다. 이것은 주변 잡음의 스펙트럼을 추정하여 입력된 음향에서 차감하는 방법으로 백색 잡음이나 자동차 실내 소리와 같이 변화가 적은 일정한 주변 잡음 (stationary noise) 에 대해서 높은 성능을 보일 수 있는 방식이다. 기본적인 알고리즘은 다음과 같다.

$$|\widehat{S}(k)| = |X(k)| - \mu(k)$$

여기서 $X(k)$ 는 입력 음향, $\mu(k)$ 는 추정된 잡음의 스펙트럼이다. 위의 수식에서 음수가 나올 때 0으로 치환할 경우 residual noise 라는 특유의 잡음이 남게 되는데, 이것을 제거하기 위하여 위의 수식을 변형하여서도 널리 사용한다.

III. 미등록 음향 거절 문제

인식기가 항시 주변의 소리를 입력받다 원하는 소리가 발생 시에만 검출해내기 위해서는 시스템의 어휘 내에 속한 소리만을 구별해내는 방법이 필요하다. 일반적인 음성 인식기에서는 음성 구간 검출 (voice activity detection) 모듈이 사람의 목소리와 그 외의 소리를 걸러내는 역할을 하고 있으며, 핵심어 검출 (keyword spotting) 시스템의 경우 등록된 어휘 이외의 미등록어 (out-of-vocabulary word) 를 걸러내기 위하여 confidence measure 등의 방법으로 인식 결과가 등록 어휘에 속하는지를 평가하였다.

음성 구간 검출은 여러 다른 소리와는 구별되는 사람의 목소리만의 특징을 찾는 것을 목적으로 하는 연구 분야이다. 기본적인 에너지 기반의 알고리즘 [2] 에서부터 음성 스펙트럼의 특징을 이용한 여러 알고리즘 [3] [4] 등이 제시되었다.

인식 결과 후 신뢰도를 측정하여 잡음인지 여부를 판별하는 방식은 핵심어 검출 기술 (keyword spotting) 분야에서 오랫동안 연구된 방식이다. 이것은 인식 결과가 맞을 때와 틀렸을 때 현저하게 다른 분포를 보이는 특징을 찾는 것을 목적으로 한다. 주로 인식 결과의 우도값 (likelihood) 과 나머지 후보들 혹은 필러 모델 (filler model) 의 비율을 측정하여 거절하는 방법이 이용된다 [5].

음성 구간 검출이나 신뢰도 측정 방식은 아직까지는 모든 종류의 잡음에서 음성만 정확히 구별하거나 어휘 이외의 단어 등을 확실히 검출할 수 있는 방법이 발견되지 않은 상태이다. 보청견 대체 인식기와 같이 음성 이외의 소리를 검출하는 것을 목표로 하는 경우에는 음성의 특징을 이용한 알고리즘은 적용할 수 없다. 이런 경우 음성의 일반적인 특징 대신 개별 음향의 특징을 이용할 필요가 있는데, 대상 음향의 특징에 대한 별도의 사전 지식이 없는 경우 스펙트럼 상에 높이 솟아오르는 봉우리 (spectral peak) 등을 이용하는 방법이 있을 수 있다.

IV. 결론

본 논문에서는 핵심어 검출기술에 기반한 시스템을 개발할 때의 두 가지 문제에 대하여 알아보았다. 음향 왜곡을 해결하기 위한 기존의 연구로는 왜곡된 음향과 인식 시스템의 학습 데이터와의 차이를 보정하는 방법과 왜곡의 영향을 덜 받는 특징 추출 및 유사도 측정 방법이 있었다. 미등록 음향의 거절을 위한 기존 연구는 인식 이전에 잡음구간과 음성 구간을 구분하는 음성 구간 검출 방법과 인식 이후에 인식 결과의 신뢰도로 대상 어휘와 그 외를 구분하는 방법이 있었다.

핵심어 검출 시스템에 있어서 인식의 전 단계에 해당하는 음성 구간 검출 방식의 경우 잡음을 잘못 받아들인 경우에는 이후의 신뢰도 측정 단계에서 한 차례 더 걸러질 가능성이 있어서 회복 가능 오류에 속한다. 반면 음성을 잘못 잡음으로 거절한 경우는 이후 단계에서 회복할 방법이 없으므로 시스템을 구현할 때에는 음성 구간 검출 단계의 문턱값 (threshold) 를 여유롭게 두어서 음향이 잘못 거절되는 일을 줄이고 잘못 들어온 잡음은 이후 신뢰도 측정 단계에서 정밀하게 걸러낼 수 있도록 구현하는 것이 바람직하다 볼 수 있다.

감사의 말

본 논문은 정통부 및 정보통신연구진흥원의 정보통신선도기반기술사업의 연구결과로 수행되었습니다.

참고문헌

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-27(2), pp.113-120, 1979

[2] L. F. Lamel, L. R. Rabiner, A. E. Rosenberg, and J. G. Wilpon, "An improved endpoint detector for isolated word recognition", *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-29(4), pp.777-785, 1981

[3] B. Wu, and K. Wang, "Robust endpoint detection algorithm based on the adaptive band-partitioning spectral entropy in adverse environments", *IEEE Transactions on Speech and Audio Processing*, vol.13, pp.762-775, 2005

[4] J. Ramirez, J. Segura, C. Benitez, A. Torre, and A. Rubio, "A new adaptive long-term spectral estimation voice activity detector", *Proc. Eurospeech*, pp. 3041-3044, 2003

[5] H. Jiang, "Confidence measures for speech recognition : A survey", Speech Communication, vol.45, pp. 455-470, 2005