잡음 환경에서의 인식 거부 성능 향상을 위한 신뢰 척도

박 정 식, 오 영 환
한국과학기술원 전자전산학부

Confidence Measure for Utterance Verification in Noisy Environments

Jeong-sik Park, Yung-hwan Oh
Department of Electrical Engineering & Computer Science
Korea Advanced Institute of Science and Technology
E-mail : {dionpark,yhoh}@speech.kaist.ac.kr

Abstract

This paper proposes a confidence measure employed for utterance verification in noisy environments. Most of conventional approaches estimate the proper threshold of confidence measure and apply the value to utterance rejection in recognition process. As such, their performance may degrade for noisy speech since the threshold can be changed in noisy environments. This paper presents further robust confidence measure based on the multi-pass confidence measure. The isolated word recognition based experimental results demonstrate that the proposed method outperforms conventional approaches as utterance verifier.

I. 서론

각종 모바일 기기의 등장과 함께 텔레매틱스 및 모바일 환경에서의 음성 인식 시스템의 신뢰성 향상을 위한 다양한 연구가 수행되었다. 특히 잡음에 의한 오인식 결과 및 미등록 어휘(Out-Of-Vocabulary (OOV))에 대한 거부 처리는 인식기의 성능 향상을 위한 필수 과정으로 적용된다[1][2].

대다수의 음성인식 시스템은 인식 단계에서 후보 어휘에 대한 순위를 매기기 위한 목적으로 확률, 즉 가정(hypotheses)에 근거하여 점수를 계산한다. 그러나 이 점수는 다른 후보에 비해 가능성이 크다는 것을 보일 뿐 가정이 올바른지 또는 틀린지에 대한 좋은 지표를 제공하는데 한계가 있다. 사용자의 입력 음성에 대하여 적절한 피드백이 제공되는 음성 인식 서비스의 경우, 인식 단계에 사용되는 사후 확률(posterior probability)이 서비스의 신뢰성과 직접적으로 연관되며, 가정에 의해 결정된 단어(혹은 문장)의 정확성이 어느 정도인지를 측정하는 발화 검증(utterance verification) 과정이 요구된다. 발화 검증을 통해 OOV, 감탄사, 잡음 음성 등 오인식의 원인이 되는 발화를 효과적으로 거부함으로써 인식 성능을 향상시킬 수 있으며, 의도하지 않은 발화가 입력되는 경우 재차 질문을 던져 정확한 발음을 유도하거나 상담원에게 자동 연결하는 등 신뢰성 있는 서비스 제공이 가능하다.

신뢰 척도는 인식 결과의 승인/거부를 판별하는 기준으로서 발화 검증의 대표적인 방법이며, 대표적인 신뢰 척도 파라미터로 인식 결과(N-best) 및 사후 확률(또는 우도)이 사용된다[3][4]. 하지만 기존의 신뢰 척도는 잡음 음성에 적용하는데 한계가 있으며, 본 연구에서는 이를 해결하기 위한 방법으로 다중 음성 인식 결과에 기반한 신뢰 척도를 제안한다.

2장에서는 기존의 신뢰 척도 및 개선점을 제시하고 3장에서 잡음 음성에 효과적인 신뢰 척도를 제안한다. 4장에서는 제안한 방법의 유효성을 검증하는 실험 결과를 살펴본 후 5장에서 결론을 맺는다.

II. 신뢰 척도

음성 인식의 결과는 주어진 훈련 모델에 대한 입력 음성의 사후 확률(우도)이 최대화되는 예측 모델로 결

정된다. 이처럼 확률에 의해 예측된 결과를 검증하기 위한 목적으로 다음과 같은 신뢰 척도 기법들이 소개되었다.

첫째, 인식 과정에 사용되는 정보를 이용하여 인식 결과의 신뢰도를 측정하는 방법이다. 발화의 길이(duration), 주파수 대역 분포 특성을 비롯하여 LSA (Latent Semantic Analysis), MI (Mutual Information) 등을 사용하는 다양한 방법이 있으나 N-best 결과 및 우도에 기반을 둔 신뢰 척도의 성능이 좋다는 연구 결과가 있었다[5][6]. 이 방법은 특히 별도의 filler model을 사용하여 OOV를 선별하는 방법에 비해서도 효율적이라고 알려져 있다[7]. 두 번째 방법은 다중 인식 시스템을 이용하는 방법으로 두 종류 이상의 인식기를 구축한 뒤 각 시스템에서 얻어진 출력(음향학적 확률 또는 우도)에 의해 인식 결과를 검증한다. 시스템 구축에 소요되는 자원 및 비용에 대한 부담이 큰 반면, 발화 검증에 있어 뛰어난 성능을 보인다[8].

N-best 및 우도 기반의 신뢰 척도로 사용되는 대표적인 방법은 다음과 같다.

$$CM = \frac{L_1}{\sum_{i=1}^{N} L_i} \quad (1), \qquad CM = L_1 - \sum_{i=1}^{N} L_i / N \quad (2)$$

두 가지 식에서 L_i는 N-best 결과 중 i번째 결과의 우도를 의미한다. 식 (1)은 인식 결과의 우도(L_1)를 N개의 우도의 합으로 정규화한 값이며 식 (2)는 N개의 우도의 평균을 pseudo-filler 모델로 사용한 것으로, 정확히 인식된 결과일수록 L_1과 다른 우도 사이의 차가 크다는 성질로부터 유도되었다. 이와 같은 CM은 정확히 인식된 결과와 그렇지 않은 결과 사이에 발생하는 차이를 통해 인식 결과의 검증에 사용된다. 이 때, 인식 결과의 승인/거부를 판별하는 임계값(threshold)이 결정되어야 하며, 잘못 거부된 발화의 비율(False Rejection rate (FRR))과 잘못 승인된 발화의 비율(False Acceptance rate(FAR))이 비슷한 값을 임계값으로 정한다. 간편하면서 효과적인 신뢰 척도로 알려진 이 방법은 미리 정해진 임계값에 의존하여 승인/거부가 결정되므로 시스템 혹은 입력 음성의 환경에 따라 임계값의 지속적인 갱신이 요구된다. 따라서, 잡음의 종류 및 세기가 심하게 바뀌는 환경에서는 인식 오류가 발생하는 문제가 있다. (이와 관련된 실험 결과를 4.2장에서 제시한다.) 미등록 어휘 및 잡음을 별도의 garbage model로 구성하여 인식 거부에 사용하는 연구도 있었으나[2], 잡음의 다양한 종류 및 세기를 고려하였을 때 정확한 발화 검증에 한계가 있다. 또한, 다중 인식 시스템에 의한 검증 방법으로서 다양한 잡음 처리 기법이 적용된 다수의 인식 시스템을 통해 정확한 인식 결과를 예측할 수 있지만 시스템 구축에 대한 부담이 따른다.

III. Multi-pass 신뢰 척도

입력 음성에 존재하는 배경 잡음의 처리는 음성 인식 시스템의 안정된 성능을 보장하는 필수 과정이며, 잡음에 강인한 특징 파라미터, 모델 보상, 입력 음성의 잡음 제거 등 다양한 방법이 대부분의 인식기에 적용되고 있다. 본 연구에서는 잡음 환경에서의 효과적인 발화 검증을 위한 방법으로 잡음 처리를 통해 잡음이 제거된 음성과 원래의 음성에 대한 인식 결과를 기반으로 신뢰 척도를 계산하는 방법을 제안한다.

3.1 다중 인식 결과 기반의 신뢰 척도

일반적으로 OOV에서 계산된 CM은 등록 어휘(In-Vocabulary (IV))의 CM과 큰 차이를 보이나, 잡음에 의해 오인식되는 IV의 경우 OOV의 CM과 비슷한 결과를 나타낸다. 따라서 임계값 측정 시 사용된 자료의 잡음 세기(SNR)와 입력 음성의 SNR이 다른 경우 등록 어휘임에도 불구하고 거부로 판별될 여지가 크다.

본 논문에서는 잡음 환경에서 발생하는 발화 검증 오류를 효과적으로 해결하기 위해 그림 1과 같은 검증 방법을 제안한다. 입력 음성과 잡음이 제거된 음성의 인식 결과(N-best 및 우도)로부터 신뢰 척도(식 (1) 또는 (2))를 각각 계산한 뒤 이들의 차이를 발화 검증에 사용한다. 제안한 방법은 오인식될 가능성이 큰 발화와 정확히 인식되는 발화 사이에 신뢰 척도의 차이가 크다는 점에 기인한다. 즉, 잡음 처리 후 오인식된 발

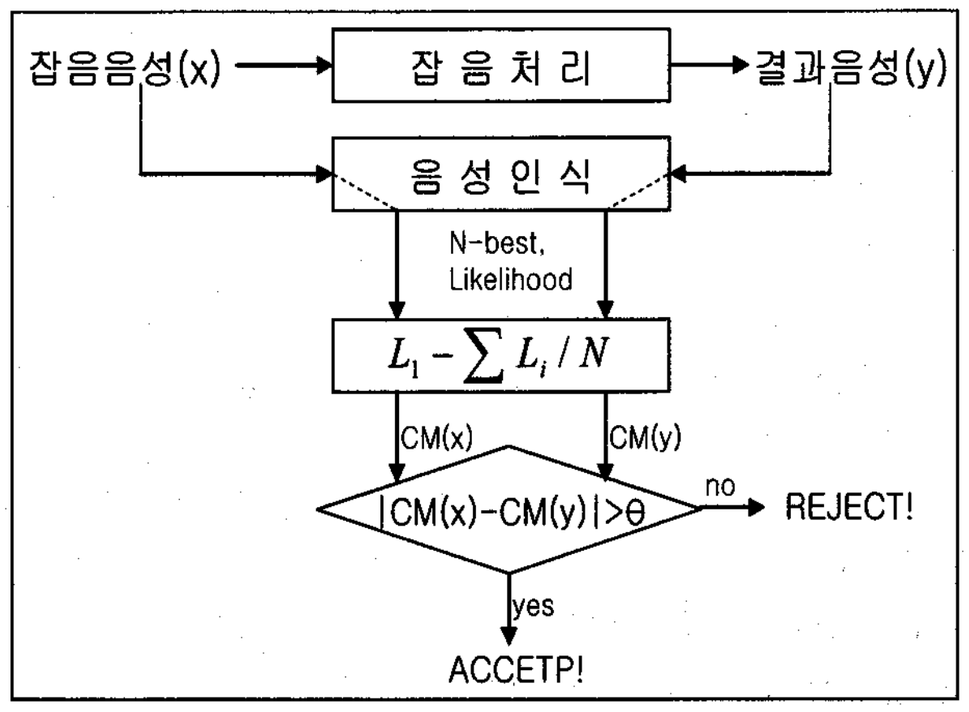


그림 1. 다중 인식 기반의 발화 검증

화는 잡음 처리 전에도 오인식될 확률이 크며, 이 경우 CM(x)와 CM(y)의 차이는 작다. 반면, 잡음이 제거된 후 정확히 인식된 발화의 경우 잡음이 제거되기 전에는 인식 결과와 무관하게 잡음에 의해 상대적으로 작은 CM을 나타내므로 CM(x)와 CM(y)의 차이가 크다. 이 같은 성질에 의해 |CM(x)-CM(y)|의 값이 작은 음성은 거부하고 차이가 큰 음성은 승인한다.

3.2 multi-pass 신뢰 척도

제안한 방법은 단일 인식 시스템을 사용하므로 2장에서 살펴 본 다중 인식 시스템 기반의 신뢰 척도에 비해 소요되는 자원 및 비용 부담이 적은 장점이 있으나 두 차례의 인식 과정으로 인해 비슷한 처리 시간이 요구된다. 또한, 실험실 환경의 음성의 경우, 잡음 처리 전후의 SNR의 차이가 크지 않으므로 신뢰 척도에 의해 잘못 거부되는 결과가 발생할 우려도 있다. 그림 2의 알고리즘에 의한 multi-pass 신뢰 척도는 이 같은 문제를 보완한다.

step 1에서는 기존의 신뢰 척도에 의해 승인/거부를 판별하며, 두 가지 임계값 즉, 확실한 승인을 검증하기 위한 γ_1과 확실한 거부의 기준을 의미하는 γ_2을 사용한다. 따라서 step 1에서 승인/거부가 결정되는 발화는 한 번의 인식 과정만 요구된다. 반면, CM(y)가 γ_1과 γ_2 사이에 존재하는 발화는 기존의 신뢰 척도로 확실하게 검증되지 않는 발화로 간주되어 두 번째 단계로서 다중 인식 결과 기반의 신뢰 척도를 수행한다. CM(x)과 CM(y)의 차를 이용하는 step 2에서는 N-best 및 우도에 의한 성질을 함께 적용한다. 즉, x (잡음 처리 전의 발화)와 y (잡음이 제거된 발화)의 인식 결과가 모두 정답인 경우는 인식 결과(N-best list 중 첫 번째)가 동일하다는 점, 그리고 y의 결과가 정답이고 x는 오답인 경우 CM(y)-CM(x)의 값이 항상 양수임을 이용한다.

[step 1]
if(CM(y) > γ_1) accept y
else if((CM(y) < γ_2) reject y

[step 2]
if(1st_best(x) = 1st_best(y) or CM(y) > CM(x))
  if(|CM(x)-CM(y)| > θ) accept y
  else reject y
else reject y

그림 2. multi-pass 신뢰 척도 알고리즘

Ⅳ. 인식실험 및 결과

4.1 실험 환경

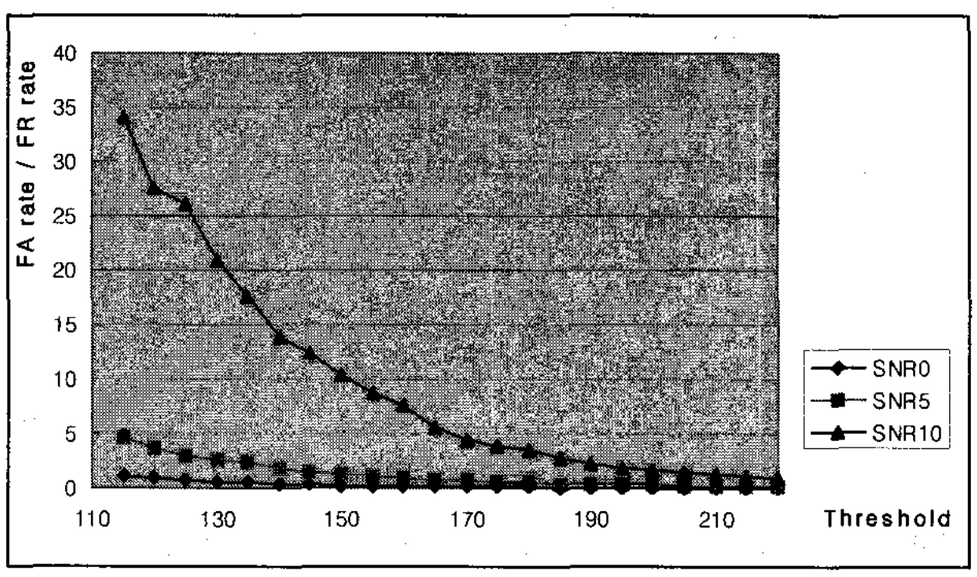
제안한 방법의 성능 평가를 위해 HTK 기반의 고립 단어 인식 실험을 수행하였다. 한국어 인식 성능 평가를 목적으로 한국어 낭독 음성 PBW (Phonetically Balanced Word) DB를 실험 자료로 사용하였다. 200개 단어에 대해 38명의 화자가 실험실 환경에서 두 차례씩 발성한 총 15200개의 발화를 학습 자료로 사용하였으며, 10명의 화자로 구성된 6000개의 발화를 평가 자료로 사용하였다. 평가 자료에는 50개의 미등록 어휘를 발성한 600개의 자료를 OOV로 포함시켰으며, 모든 평가 자료에 SNR 0에서 15 사이의 배경 잡음을 추가하였다. 실험에 사용한 잡음의 종류는 NoiseX-92에 포함된 백색 잡음과 군중, 공장 잡음이다. 13차 MFCC (에너지 포함), 차분, 가속 MFCC로 구성된 39차 MFCC를 특징 파라미터로 사용하였으며, 각 단어마다 3 state로 구성된 트라이폰 HMM 모델을 구성하였다.

4.2 실험 결과

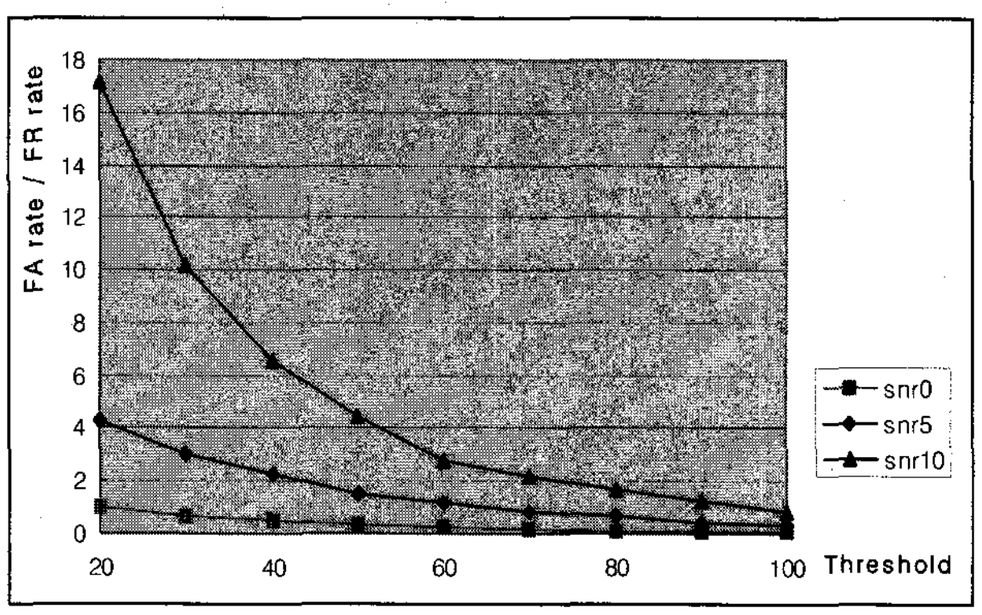
본 연구에서는 multi-pass 신뢰 척도 및 기존의 방법이 잡음 세기로 인해 받는 영향을 비교하기 위한 실험과 제안한 발화 검증 방법의 성능을 평가하기 위한 실험을 수행하였다.

그림 3은 기존의 신뢰 척도(a)와 제안한 방법(b)에 대하여 SNR 및 임계값에 따른 False Acceptance Rate (FAR)과 False Rejection Rate (FRR)의 비를 도식화한 것이다. 두 경우 모두 pseudo-filler 모델 기반의 신뢰 척도(식 (2))를 사용하였으며 잡음 처리 방법으로 스펙트럼 차감법(Spectral Subtraction)을 적용하였다. 일반적으로 SNR이 낮아질수록 적정 임계값[1]이 작아지는 경향을 보이며, 임계값의 변화 범위가 각 신뢰 척도마다 다르므로, SNR이 0인 환경에서의 적정 임계값과 SNR이 10인 환경에서의 적정 임계값 사이에서 FAR/FRR의 추이를 관찰하였다. 가장 이상적인 신뢰 척도는 모든 SNR에 대하여 임계값이 변함에 따라 FAR/FRR이 동일하게 변화하는 경우이다. 기존의 신뢰 척도의 경우, SNR 0에서의 적정 임계값을 SNR 10인 잡음 환경에 적용하여 발화 검증을 수행할 때, FAR/FRR의 값이 35까지 상승하는 반면, 제안한 신뢰 척도에서는 17 정도에 머무르는 것을 확인하였다.

1) 적정 임계값은 FAR/FRR이 1에 가까워질 때의 threshold이다.



(a) 기존의 신뢰 척도



(b) 제안한 신뢰 척도

그림 3. SNR 및 임계값에 따른 FAR/FRR의 변화

이 결과는 제안한 방법이 기존의 방법에 비해 입력 음성의 잡음 세기의 영향을 적게 받음을 의미한다.

배경 잡음 및 SNR이 다양하게 포함된 전체 실험 자료를 대상으로 False Acceptance rate과 False Rejection rate을 조사하여 그림 4와 같은 DET (Detection Error Trade-off) 결과를 확인하였다. 곡선이 원점에 더 가깝게 인접한 multi-pass 신뢰 척도가 기존의 방법에 비해 향상된 척도임을 나타낸다. FAR과 FRR이 동일한 지점에서 ERR (Equal Error Rate)을 조사한 결과, 제안한 방법에서는 56.5%, 기존의 방법은 59.3%의 성능을 보였다.

V. 결론

본 논문에서는 잡음 환경에 효과적인 발화 검증 방법으로서 multi-pass 신뢰 척도를 제안하였다. 잡음이 처리되기 전의 음성과 잡음이 제거된 음성 사이에서 발생하는 신뢰 척도의 차이를 신뢰 척도로 이용함으로써 잡음 환경에서의 인식 거부 성능이 향상됨을 확인하였다. 한국어 음성을 대상으로 고립 단어 인식 실험을 수행한 결과 기존의 방법에 비해 약 3%의 EER 성능 향상을 보였다. 향후, 다양한 신뢰 척도를 제안한

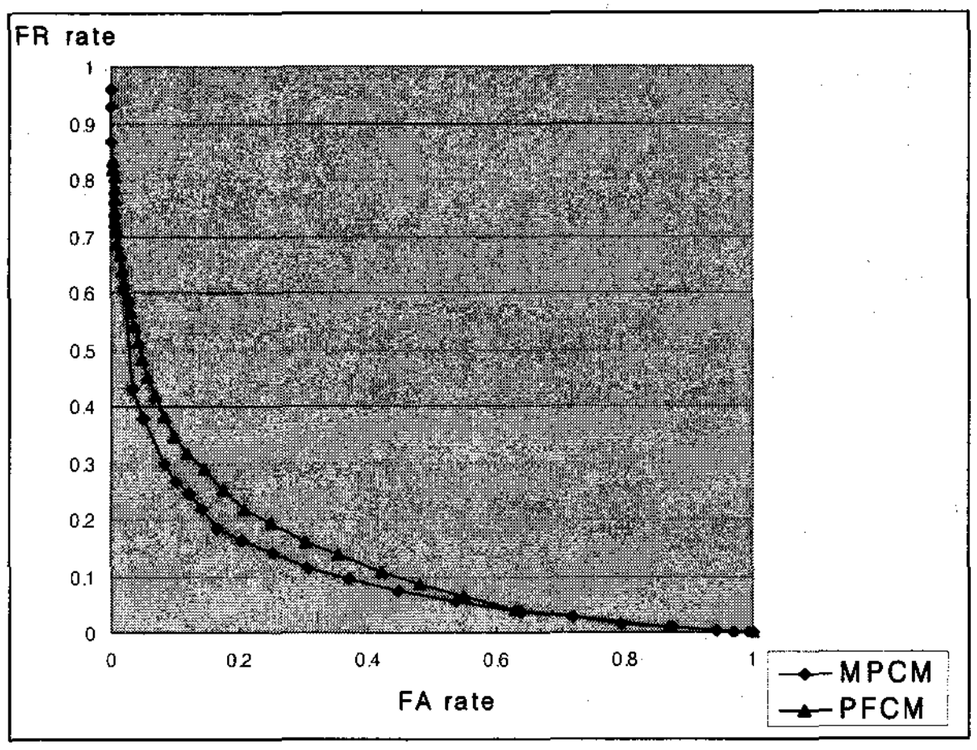


그림 4. Multi-pass 신뢰 척도(MPCM)와 Pseudo-filler 모델 기반의 신뢰 척도(PFCM)의 DET 분포

방법에 적용함으로써 추가적인 성능 향상이 검증되어야 하며, Aurora 등 외국어 DB를 사용하여 성능 평가를 수행할 필요가 있다.

참고문헌

[1] D.Wu, Tanaka M., "A robust speech detection algorithms for speech activated hands-free applications," *Proc. ICASSP*, pp.2407-2410, 1999

[2] H. Jiang, "Confidence measures for speech recognition : a survey," *Speech Communication*, vol.45, no.4, pp.455-470, Apr.2005

[3] Hernandez-Abrego, G., "Robust and efficient confidence measure for isolated command recognition," *Proc. ASRU*, pp.449-452, Dec.2001

[4] Yuewen Fu; Limin Du, "Combination of multiple predictors to improve confidence measure based on local posterior probabilities," *Proc. ICASSP*, pp.93-96, 2005

[5] Gang G.; Chao H., "A comparative study on various confidence measures in large vocabulary speech recognition," *Proc. ISCSLP*, pp.9-12, 2004

[6] J. Dolfing, A. Wendemuth, "Combination of confidence measures in isolated word recognition," *Proc. ICSLP*, pp.3237-3240, 1998

[7] Greenland, G., Wong, W, "Improving utterance verification using additional confidence measures in isolated speech recognition interfaces," *Proc. ICASSP*, pp.81-84, 2005

[8] Vincent V., "Confidence scoring and rejection using multi-pass speech recognition," *Proc. Eurospeech*, pp.3133-3136, 2005