광대역 LSP 계수의 패턴 추론을 위한 연판정 기반 개선된 조합 매핑 설계

전종근*
*LIG넥스원

e-mail:jonggeun.jeon@lignex1.com

Designing on improved combined mapping based on soft-decision for wideband LSP coefficients pattern estimation

Jong-geun Jeon*
*Avionics Lab., LIG Nex1

요 약

본 논문은 인공 대역 확장에서 스펙트럼 포락선 확장 시 발생하는 스펙트럼 왜곡을 줄이는 개선된 조합 매핑(Improved combined mapping) 알고리즘을 제안한다. 벡터양자화를 기반으로 하는 코드북 매핑(Codebook mapping)과 스펙트럼 포락선(Spectrum Envelope)의 선형 의존도를 이용한 선형 매핑(Linear mapping)을 사용하여 각각 확장된 광대역 LSP(Line Spectrum Pair)를 추론하고, 연판정(Soft-decision)을 통해 최적화된 LSP를 추론한다. 제안된 알고리즘으로 합성된 음성신호의 스펙트럼 왜곡(Spectrum Distortion)이 기존 조합매핑으로 얻은 음성 신호의 스펙트럼 왜곡보다 더 적은 왜곡을 갖는 결과를 나타내었다.

1. 서론

음질 개선을 위해 사용되는 알고리즘 중 하나인 인공 대역 확장 알고리즘은 소스-필터 모델에 기반으로 하는 알고리즘이다. 인공대역 확장 시, 음성은 LP분석 필터를 통해 여기신호 및 스펙트럼 포락선 성분을 각각 추출하여 각각의 확장 알고리즘을 통해 확장 후 재합성을 하게 되고, 기존 입력신호에 비해 개선된 품질의 음성을 출력하게 되는 것이다.

한편 스펙트럼 확장을 위해 스펙트럼 성분을 갖는 LSP(Line Spectrum Pair)의 패턴을 이용하여, 확장된 LSP를 얻기 위해 기존에 제안되었던 대표적인 알고리즘은 코드북 매핑, HMM, GMM 등이 있다. [1][2] 코드북 매핑(codebook mapping)방식은 저대역 스펙트럼 포락선과 고대역 스펙트럼 포락선의 LSP 패턴의 상관관계를 이용하여 입력신호의 스펙트럼과 매핑되는 코드북과, 이에 대응되는 스펙트럼 코드북으로 확장된 스펙트럼을 구성한다. HMM(Hidden Markov Model) 방식은 입력 음성 프레임으로부터 출력된 코드벡터가 HMM을 위한 관찰값으로 사용되고, HMM의 각 스테이트는 확장대역의 스펙트럼을 정의하는 엔트리 값으로 사용된다. GMM(Gaussian Mixture Model) 기반 스펙트럼 포락선 확장은 스펙트럼 코드벡터의 확률모델에 기반하여 스펙트럼 대역폭을 확장하는 방법이다. 선형 매핑(Linear mapping)은 입출력 신호의 스펙트럼 포락선 선형 의존도(linear dependency)를 이용한 직관적이면서 아주 간단한 방법이다.

본 논문에서는 앞서 언급한 여러 가지 LSP 확장 기법 중에서 음질의 스펙트럼 왜곡(Spectrum Distortion) 측면에서 기존 알고리즘보다 개선된 조합 매핑(improved combined mapping)방식을 제안한다. 또한 입력 신호는 협대역 음성신호(narrowband speech)를 이용하여 협대역 LSP를 추출하였고, 제안된 알고리즘을 이용하여 스펙트럼 왜곡이 적은 광대역 LSP를 추론하였다.

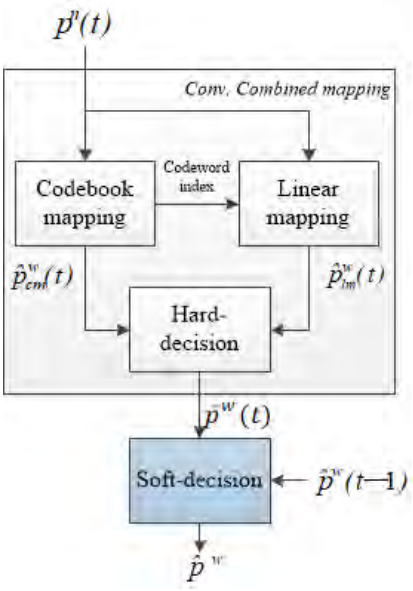
2. 개선된 조합 매핑 방식 알고리즘

서론에서 언급한 코드북 매핑 방식은 성능은 우수하다는 장점이 있지만 계산량이 많다는 것이 단점이며, 선형 매핑은 설계된 매트리스를 이용하여 계산량이 작은 장점이 있지만 성능이 나쁘다는 단점이 있다. 그래서 각 방식들의 장점을 개선하기 위해 조합 매핑 방식이 제안되었다.[3] 조합 매핑 방식은 코드북 매핑을 통해 얻은 코드워드 인덱스를 이용하여 선형 매핑에 필요한 매트리스를 선택하고, 각 매핑을 통해 확장된 광대역 LSP는 입력신호의 협대역 LSP와의 유클리디언 거리(Euclidean Distance)를 비교하여 최소가 되는 값으로 사용하는 경판정(hard-decision) 함으로써 최종 광대역 LSP를 출력한다.

하지만 이런 알고리즘을 통한 스펙트럼 추정 예측방식은 실제 음성의 광대역 스펙트럼과의 거리가 먼 포락선으로 수행할 가능성도 있다. 이럴 경우 유성음 구간에서 프

레임간 음성 스펙트럼의 상관관계 특성과는 전혀 다른 스펙트럼이 출력되고, 이는 곧 음성 품질의 성능 저하를 일으킬 수 있다. 이처럼 경판정으로부터 간혹 발생하는 음성 품질 저하 증상을 막기 위해 연판정 방식을 본 논문에서 적용하였다.

그림 1은 기존 조합 방식을 기반으로 연판정 방식을 적용한 개선된 조합 매핑 구조이다.



(그림 1) 연판정을 통한 개선된 조합 매핑 구조

VAD(Voice Activity Detector)를 이용하여 유성음과 무성음 구간을 판단하고,[4] 유성음 구간에서 추출된 협대역 LSP $p^n(t)$를 이용하여 본 논문에서 제안된 매핑구조의 입력으로 사용하였다. 기존 조합매핑 방식에서 코드북 매핑으로 추출된 $\hat{p}^w_{cm}(t)$과 선형매핑을 통해 추출된 $\hat{p}^w_{lm}(t)$는 협대역 입력신호의 $p^n(t)$와 유클리디언 거리를 계산하여 거리가 더 낮은 $\bar{p}^w(t)$를 얻게 된다. 여기서 이전 프레임에서 결정된 광대역 LSP $\hat{p}^w(t-1)$와 조합 매핑에서 선택된 $\bar{p}^w(t)$에 식(1)과 같은 가중치를 주어 최종 광대역 LSP $\hat{p}^w(t)$를 추출하게 된다. 여기서 α는 가중치 벡터(weight vecor)이다.

$$\hat{p}^w(t)=\alpha\bar{p}^w(t)+(1-\alpha)\hat{p}^w(t-1) \quad \text{for} \quad 0\le\alpha\le 1 \qquad (1)$$

표1은 최적의 α의 값을 결정하기 위해, α의 변화에 따른 $\hat{p}^w(t)$의 로그 스펙트럼 왜곡값 비교를 나타낸 표이다. 왜곡값 비교를 위해 약 5분 정도 길이의 협대역 및 광대역 음성 샘플을 사용하였고, 광대역 원음성의 3400 ~ 7000Hz 고대역 스펙트럼과 추론을 통해 획득된 광대역 음성의 고대역 로그 스펙트럼 왜곡을 비교하였다.

<표 1> α 값의 변화에 따른 고대역 로그 스펙트럼 왜곡 비교

α	0.5	0,6	0.7	0.8
LSD(dB)	4.11	3.89	3.97	4.09

표1에서 볼 수 있듯이, α가 0.6일 때 가장 작은 스펙트럼 왜곡을 나타내었다.

3. 실험결과

본 논문에서 제안된 연판정 기반 개선된 조합 매핑 방식의 개선결과를 비교하기 위해, 약 5분정도 길이의 협대역 및 광대역 음성 샘플 2세트를 사용하였다. 각각의 음성 세트는 다수의 남성과 여성의 목소리가 녹음되어 있으며, 이를 이용하여 실험결과를 도출하였다. 코드북 매핑에서 사용되는 코드북 사이즈는 128, 협대역 LSP의 차수는 10차, 광대역 LSP의 차수는 20차로 설정하였다. 선형매핑에서 사용되는 매트릭스는 6개로 설정하였다. 식(2)는 광대역 원음성과 LSP 추론을 통해 합성된 광대역 음성의 고대역 성분(3400 ~ 7000Hz)의 로그 스펙트럼 왜곡(Log Spectrum Distortion) 정의이다.

$$LSD(dB)=\sqrt{\frac{1}{7000-3400}\int_{3400}^{7000}(10\log_{10}P(f)-10\log_{10}P'(f))^2df} \qquad (2)$$

식(2)에서 $P(f)$는 광대역 LSP를 AR계수 a_w로 변환하여 a_w의해 계산되는 파워 스펙트럼(Power Spectrum)이며, $P'(f)$는 개선된 조합 매핑을 통해 얻은 광대역 LSP를 AR계수로 변환하여 얻은 파워 스펙트럼이다. $P(f)$는 식(3)과 같이 정의된다.

$$P(f)=1/\left|1-\sum_{k=1}^{20}a_w^k e^{-jk2\pi f}\right|^2 \qquad (3)$$

표2는 음성 샘플에 따른 기존 조합 매핑의 로그 스펙트럼 왜곡값과 제안된 조합 매핑의 로그 스펙트럼 왜곡값을 비교하였다.

<표 2> 로그 스펙트럼 왜곡(LSD) 값 실험 결과

방법 / LSD(dB)	기존 조합 매핑	제안된 조합 매핑
남성 샘플	3.75	2.73
여성 샘플	4.47	3.89

표2에서 볼 수 있듯이, 음성 샘플의 종류에 상관없이 기존 조합 매핑방식보다 연판정을 이용한 제안된 조합 매핑 방식이 더 낮은 로그 스펙트럼 왜곡값을 가지는 걸 확인

할 수 있다.

4. 결론

본 논문에서는 기존의 경판정 기반 조합 매핑방식을 사용할 경우, 갑작스런 광대역 스펙트럼 포락선의 변화로 인해 발생하는 음성 왜곡을 줄이기 위해 이전 프레임에서 얻은 광대역 LSP를 이용하여, 기존 조합 매핑방식에 연판정을 도입하여 음성의 스펙트럼 왜곡을 줄이는 방법을 제안하였다. 실험 결과 기존 조합 매핑방식에 비해 본 논문에서 제안된 연판정 기반 조합 매핑방식이 더 낮은 음성의 스펙트럼 왜곡값을 가졌다.

참고문헌

[1] Peter Jax and Peter Vary, "On Artificial Bandwidth Extension of Telephone Speech," Signal processing, vol. 96, no. 6, pp 1276-1306, 2006.

[2] J. Y. Park and H. S. Kim, " Narrowband to wideband conversion of speech using GMM based transformation," in Proc. of ICASSP, vol. 3, pp. 1843-1864, 2000.

[3] Jonggeun Jeon, Yaxing Li, Sangwon Kang, Kihyun Choo, Eunmi Oh, and Hosang Sung, "Robust artificial bandwidth extension technique using enhanced parameter estimation," in Proc. of the Audio Engineering Society Convention (AES), Los Angeles, USA, 2014.

[4] Ramírez J., Górriz J.M., Segura J.C. "Voice Activity Detection Fundamentals and Speech Recognition System Robustness." In: Grimm M., Kroschel K., editors. Robust Speech Recognition and Understanding. I-Tech, pp. 1 - 22, Vienna, Austria: 2007.