

Spike train decoding에 기반한  
인공와우 어음처리방식 성능평가

연세대학교 대학원

의 공 학 과

김 두 희

Spike train decoding에 기반한  
인공와우 어음처리방식 성능평가

지도 김 경 환 교수

이 논문을 석사 학위논문으로 제출함


2007년 6월 일


연세대학교 대학원


의 공 학 과

김 두 희

# 김두희의 석사 학위논문을 인준함

심사위원 김 경 환 

심사위원 전 법 먼 

심사위원 이 창 환 

연세대학교 대학원

2007년 6월 일

## 감사의 글

졸업을 하며 시간이 참 빠르게 흘러간다는 것을 새삼 느낍니다. 입학할 때 그렸던 제 모습을 되돌아보며 아직도 갈 길이 멀었다는 생각이 듭니다. 부족한 제자를 믿고 많은 가르침을 주신 김경환 교수님께 감사드립니다. 자부심을 가지고 연구에 임할 수 있도록 도와주시고 재미있는 연구를 할 수 있는 기회를 주셨습니다.

연구와 강의로 바쁘신 가운데 부심을 맡아주신 김법민 교수님, 임창환 교수님께도 감사의 말씀을 드립니다. 또한 학부과정 때부터 열정적으로 강의를 해주신 윤형로 교수님, 이윤선 교수님, 이경중 교수님, 김동윤 교수님, 정병조 교수님, 김지현 교수님, 이상우 교수님, 임창환 교수님, 윤대성 교수님, 서종범 교수님, 이용흠 교수님, 권기훈 교수님, 임도형 교수님, 박성빈 교수님께 감사드립니다.

신경 공학 연구실 선후배님들께 감사드립니다. 훌륭한 요리솜씨로 랩동생들을 챙겨준 자현누나, 먼저 졸업해 열심히 연구를 지속하고 있는 진호형, 성실하고 성격 좋은 상백형, 많은 웃음을 준 콤비 진, 성진에게 고마움을 전합니다.

대학원 생활에 많은 조언과 먹을 것으로 힘을 준 승록형, 기윤형, 병관형, 은근형, 진욱형, 용현형, 운회형 같이 밥을 먹으며 즐거운 담소를 나누었던 멋진 동기 준섭, 인성, 봉렬, 진수, 연식, 동희, 명현, 현우에게도 고마움을 전합니다. 백운관에 있는 몇 안되는 의공학과 대학원생으로 함께 한학기를 보낸 희제형, 영진, 한정님에게도 감사합니다.

대학생활을 통해 얻은 친구들 태훈1, 재용, 선진, 정진에게도 감사합니다. 이들과 청년기의 고민을 공유하고 즐거움을 함께하며 쌓은 우정은 저의 보배입니다.

학교에서 답답할 때면 만나 담소를 나누고 함께 여행도 자주 다닌 고등학교 친구들에게도 고마움을 전합니다. 공주의 명의 오성, 대학원 생활의 비애를 공유한 태훈2, 치열한 사회생활에 노출된 희만. 시험 준비로 바쁜 정흠이, 이들이 있었기에 즐거움과 활력을 충전할 수 있었습니다. 또한 스승의 날과 명절 전후로 같이 선생님을 찾아뵙는 준하, 혁, 종준, 윤석, 현수 그리고 제게 많은 정성을 쏟아주시

고 참된 스승의 모습을 보여주신 채수선 선생님께도 감사의 말씀을 드리고 싶습니다.

신앙적인 고민을 나누고 즐거움도 함께 했던 세희형, 은호형, 상덕형 에게도 감사드립니다. 순수한 열정을 가지고 성실한 생활을 유지해가는 모습으로 제게 자극을 주고 부족한 저와 진지하게 신앙적인 토론을 했던 것 잊지 못할 것입니다.

멀리 있지만 항상 저를 생각하고 아껴주신 부모님의 은혜에 감사드립니다. 아직도 어린아이를 걱정하듯 마음고생을 많이 하시는 부모님께 성숙한 모습을 보여드리고 싶은 아들이 되겠습니다. 떨어져 있어 많은 대화를 나누지도 못했고 함께한 시간이 적었지만 형을 믿고 따라준 동생에게 미안함과 고마움을 전합니다. 그리고 저의 졸업을 기뻐하셨을 작년에 고인이 되신 큰아버지께도 감사드립니다. 집안의 어른으로서 보여주신 강인함과 조카를 사랑해주신 마음은 영원히 잊지 못할 것입니다. 어려서부터 은근히 의지가 된 찬희형, 연락할 때마다 밥 사준다는 은희에게도 고마움을 전합니다. 함께 커오며 혈육의 정을 다진 그들은 저의 큰 자산입니다.

지금까지 저를 인도하신 하나님께 감사와 영광을 돌리며 여기에 적지 못하였지만 저를 불심양면으로 응원해주신 많은 분들께도 고마움을 전하며 글을 마칩니다.

2007년 7월

김두희 올림

# 차 례

그림 차례 .....	iii
표 차례 .....	v
국문 요약 .....	vi
제1장 서론 .....	1
1.1 연구배경 및 현황 .....	1
1.2 연구내용 및 범위 .....	3
제2장 방법 .....	6
2.1 Spike train decoding .....	6
2.2 어음처리방식 성능평가방법 .....	7
2.2.1 포먼트 디코딩을 통한 어음처리방식 성능평가 .....	7
2.2.2 음성시작점 정보 전달특성평가 .....	10
2.3 어음처리방식의 종류 .....	12
2.3.1 시불변-선형 필터뱅크 기반 어음처리방식 .....	12
2.3.2 시변-비선형 필터뱅크 기반 어음처리방식 .....	12
2.3.3 적응효과 .....	14
2.4 확률적 청신경 모델 .....	15
2.5 시뮬레이션을 위한 음성제작방법 .....	17
2.5.1 포먼트 디코딩을 위한 음성제작 .....	17
2.5.2 음성시작점 정보 전달특성 평가를 위한 음성제작 .....	17
제3장 결과 .....	18
3.1 포먼트 디코딩을 통한 어음처리방식 성능평가 .....	18

3.1.1 포먼트 디코딩에 이용된 상세 파라미터 .....	18
3.1.2 포먼트 궤적의 참값과 추정값 간 비교 .....	18
3.1.3 어음처리방식 간 평균 포먼트 디코딩 성능비교 .....	21
3.1.4 여러 가지 잡음 하에서 평균 포먼트 디코딩 성능비교 .....	25
3.2. 적응효과 적용 전후의 음성시작점 정보 전달특성평가 .....	27
3.2.1 음성시작점 정보 전달특성평가에 이용된 상세 파라미터 .....	27
3.2.2 적응효과 적용 전후의 음성시작점 정보 전달특성 비교 .....	28
제4장 고찰 및 결론 .....	33
4.1 Spike train decoding에 기반한 새로운 어음처리방식 평가방법의 제안	33
4.2 제안된 평가방법을 이용한 모의실험 결과 및 고찰 .....	33
4.2.1 포먼트 디코딩 .....	33
4.2.2 음성시작점 정보 전달특성 평가 .....	35
4.3 제안된 평가방법의 유용성 .....	36
참고 문헌 .....	37
영문 요약 .....	40

## 그림 차례

그림 1	인공와우의 구조 ..... (a) 인공와우 이식도, (b) 인공와우 구조 블록도	2
그림 2.1	Spike train decoding 기반 포먼트 디코딩방법 .....	8
그림 2.2	최적선형필터를 이용한 spike train decoding .....	9
그림 2.3	Spike train decoding 기반 음성시작점 정보 전달특성 평가방법	11
그림 2.4	선형분류기를 이용한 spike train decoding .....	12
그림 2.5	어음처리방식에 기반한 전기자극펄스생성 ..... (a) 어음처리방식((b) 또는 (c))에 기반한 채널별 전기자극펄스 생성 과정 블록도, (b) 시불변-선형 필터뱅크 기반 어음처리방 식, (c) 시변-비선형 필터뱅크 기반 어음처리방식	13
그림 2.6	IHC/AN 간 화학적 시냅스에 기원하는 적응효과 .....	15
그림 2.7	확률적 청신경모델 .....	16
그림 3.1	포먼트 궤적의 참값 (실선)과 추정값 (점선) 비교 ..... (a)~(d)는 시불변-선형 필터뱅크 기반 어음처리방식에 대한 포먼 트 궤적. (e)~(h)는 시변-비선형 필터뱅크 기반 어음처리방식에 대한 포먼트 궤적. (a),(e)는 잡음이 없는 경우, (b),(f)는 5 dB SNR WGN 하, (c),(g)는 0 dB SNR WGN 하, (d),(h)는 5 dB SNR SSN 하	20
그림 3.2	어음처리방식 간 평균 포먼트 디코딩 성능비교 (자극율이 800 pps일 경우) ..... (a)는 잡음이 없는 경우, (b)는 5 dB SNR 백색잡음 하, (c)는 0 dB SNR 백색잡음 하, (d)는 5 dB SNR 음성형태 잡음 하	22
그림 3.3	어음처리방식 간 평균 포먼트 디코딩 성능비교 (자극율이 2000 pps일 경우) ..... (a)는 잡음이 없는 경우, (b)는 5 dB SNR 백색잡음 하, (c)는 0 dB SNR 백색잡음 하, (d)는 5 dB SNR 음성형태 잡음 하	23
그림 3.4	어음처리방식 간 평균 포먼트 디코딩 성능비교 (자극율이 5000 pps일 경우) .....	24



(a)는 잡음이 없는 경우, (b)는 5 dB SNR 백색잡음 하, (c)는 0 dB SNR 백색잡음 하, (d)는 5 dB SNR 음성형태 잡음 하

그림 3.5	여러 가지 종류 잡음 하의 평균 포먼트 디코딩 성능변화 (자극율이 800 pps일 경우) ..... (a)는 시불변-선형 필터뱅크 기반 어음처리방식, (b)는 시변-비선형 필터뱅크 기반 어음처리방식에 대한 평균 포먼트 디코딩 성능	26
그림 3.6	여러 가지 종류 잡음 하의 평균 포먼트 디코딩 성능변화 (자극율이 2000 pps일 경우) ..... (a)는 시불변-선형 필터뱅크 기반 어음처리방식, (b)는 시변-비선형 필터뱅크 기반 어음처리방식에 대한 평균 포먼트 디코딩 성능	26
그림 3.7	여러 가지 종류 잡음 하의 평균 포먼트 디코딩 성능변화 (자극율이 5000 pps일 경우) ..... (a)는 시불변-선형 필터뱅크 기반 어음처리방식, (b)는 시변-비선형 필터뱅크 기반 어음처리방식에 대한 평균 포먼트 디코딩 성능	27
그림 3.8	음성 /pa/를 이용한 경우 어음처리방식에 적응효과 적용 전후의 음성시작점 정보 전달특성 비교 ..... (a),(b)는 시불변-선형 필터뱅크 기반 어음처리방식에 적응효과 적용 전후 비교, (c),(d)는 시변-비선형 필터뱅크 기반 어음처리방식에 적응효과 적용 전후 비교, (a),(c)는 백색잡음 (WGN) 하, (b),(d)는 음성형태잡음 (SSN) 하	29
그림 3.9	음성 /za/를 이용한 경우 어음처리방식에 적응효과 적용 전후의 음성시작점 정보 전달특성 비교 ..... (a),(b)는 시불변-선형 필터뱅크 기반 어음처리방식에 적응효과 적용 전후 비교, (c),(d)는 시변-비선형 필터뱅크 기반 어음처리방식에 적응효과 적용 전후 비교, (a),(c)는 백색잡음 (WGN) 하, (b),(d)는 음성형태잡음 (SSN) 하	30

## 표 차례

표 3.1	음성시작점 정보 전달특성 (잡음이 없는 경우와 백색잡음이 하의 경우) .....	31
표 3.2	음성시작점 정보 전달특성 (음성형태 잡음이 하의 경우) .....	32

## 국 문 요 약

### *Spike train decoding*에 기반한 인공와우 어음처리방식 성능평가

인공와우장치는 소리정보를 담은 전기자극펄스를 이용하여 청신경을 직접 자극함으로써 청력을 회복시켜준다. 인공와우장치 구성부분 중 어음처리기는 입력되는 소리정보를 지정된 어음처리방식에 기반하여 전기자극펄스 형태로 바꾸어 청신경을 자극하므로 어음처리방식의 성능평가는 필수적이다. 인공와우의 시술 및 재활훈련 등을 통한 어음처리방식의 평가는 비용, 시간 그리고 시술대상이 사람이라는 점 등 제한이 많으므로 대부분의 연구에서 어음처리방식의 평가방법으로 acoustic simulation을 이용하고 있다. 그러나 인공와우 이식과 재활훈련 이후의 성능을 예측하기에는 한계점이 있으며 어음처리방식의 실제적인 성능평가를 위해서는 다각적인 방법을 이용하여야 한다. 본 연구에서는 새로운 어음처리방식 평가방법으로 spike train decoding에 기반한 평가법을 제안한다. 이 방법을 이용하여 인공와우 어음처리방식의 성능평가에 관한 연구를 수행하였다.

현재 널리 이용되고 있는 어음처리기는 달팽이관 (cochlear)의 tonotopy특성을 모방하여 입력음성을 주파수 성분으로 분해하는 주파수 분석부와 이로부터 전기자극펄스의 크기를 결정하는 포락선 추출부로 나눌 수 있다. 본 연구에서는 일반적으로 사용되고 있는 시불변-선형 필터뱅크 기반 어음처리방식, 잡음하에서 강인한 음성정보 전달특성을 갖도록 주파수 분석부가 수정된 시변-비선형 필터뱅크 기반 어음처리방식, 음성시작점 정보 전달특성을 개선하기 위하여 포락선 추출부가 수정된 어음처리방식의 성능평가를 수행하였다.

우선 시불변-선형 필터뱅크 기반 어음처리방식과 시변-비선형 필터뱅크 기반 어음처리방식을 이용할 경우 음성정보의 청신경 응답 인코딩특성을 관찰하였다. 두 어음처리방식을 spike train decoding에 기반한 평가방법을 이용하여 비교하였다.

실험결과 시변-비선형 필터뱅크에 기반한 어음처리방식을 채택할 경우 선형 필터뱅크에 의한 경우보다 청신경 응답에 인코딩된 음성정보를 충실히 복원할 수 있었고 잡음레벨이 강할수록 특히 음성과 유사한 형태의 잡음 하에서 잡음에 강인한 특성을 보였다.

다음으로 어음처리방식에 적응효과를 추가하여 포락선 추출특성을 개선한 어음처리방식을 이용할 경우 음성의 시작점에 대한 정보 전달이 효과적으로 이루어지는지 관찰하였다. 정상인의 경우 내측유모세포와 청신경 사이의 화학적 시냅스에 의한 적응효과가 발생하지만 청신경을 전기적으로 자극하여 소리정보를 전달하는 인공와우의 경우 이를 기대할 수 없다. 그러므로 어음처리방식에 이를 추가하여야 한다. 어음처리방식에 적응효과를 적용 전후 음성시작점 정보 전달특성을 Spike train decoding에 기반한 방법을 이용하여 평가하였다. 실험결과 어음처리방식에 적응효과를 적용함으로써 음성시작점 정보 전달특성이 개선됨을 알 수 있었다.

---

핵심되는 말 : 인공와우, 어음처리방식, 시불변-선형 필터뱅크, 시변-비선형 필터뱅크, 적응효과, spike train decoding

# 제 1 장 서론

## 1.1 연구배경 및 현황

세계보건기구에 의하면 청각장애는 55세 이상 노인의 삶의 질을 떨어뜨리는 5대 요소에 속하며, 전 세계적으로 약 7억 5천만 명 정도가 청각장애자로 분류된다 [1]. 특히 이중 2백만 명은 고도의 청각장애를 가진 이들로 청력회복을 위해서는 인공와우 (cochlear implants) 시술의 대상자가 될 수 있다. 청각장애에는 전음성 청각장애 (conductive hearing loss)와 감각신경성 청각장애 (sensorineural hearing loss)가 있다 [1,2]. 전자의 경우 단순외상, 귀경화증 (otosclerosis) 등에 의해 소리 전달로의 장애가 생긴 것이 원인으로 단순히 소리를 증폭하여 주는 보청기의 사용으로도 청력회복이 가능하다. 후자는 잡음에 과도하게 노출되거나 고령, 수막염 (meningitis) 등에 의해 와우각 내 유모세포 손실이나 신경통로 혹은 뇌의 청각담당 영역의 손상 등이 발생한 경우이다 [3].

외이를 통하여 들어온 소리는 고막을 진동시키고 이는 중이의 이소골 (stapes)을 통하여 증폭되고 달팽이관 (cochlear) 내 림프액에 전달되어 유모세포 (hair cell)를 움직인다. 유모세포가 움직임에 따라 청신경절에서 활동전위 (action potential)가 발생하여 뇌로 전달되어 소리를 인지할 수 있게 된다 [3]. 감각신경성 청각장애의 경우 유모세포의 변형이 일어나지 않으므로 보청기 등의 청각보철장치를 이용하여 소리를 증폭시키더라도 청신경의 활동전위가 발생되지 않기 때문에 청각정보가 뇌로 전달되지 않는다. 그러므로 청신경에 직접 전기자극을 가하여 청력을 회복시키는 인공와우에 의한 청각회복의 수혜자가 될 수 있다 [4,5].

인공와우 장치는 청신경에 전기자극을 인가함으로써 청력을 회복시켜주는 장치로 마이크로폰, 신호처리부(어음처리기), 신호 송/수신부, 전극으로 구성된다 [4,5]. 소리가 마이크로폰을 통해 입력되면 신호처리부에서 지정된 어음처리방식에 의해 전기자극펄스 형태로 바뀌어 달팽이관 내에 삽입된 전극으로 전달된다 (그림 1) [4,5]. 어음처리기의 어음처리방식은 달팽이관의 tonotopy 특성을 모방하여 입력되는 소리정보를 주파수 성분으로 분해하는 주파수 분석부와 이로부터 전기자극펄

스의 크기를 결정하는 포락선 (envelope) 추출부로 구성된다 (그림 1). 소리에 담긴 정보가 얼마나 충실하게 청신경계로 전달되는가는 전기자극펄스를 생성하는 어음처리에 달려있으므로 이에 대한 성능평가는 필수적이다 [4-7,21,22]. 최종적으로는 인공와우 시술자를 대상으로 어음처리의 성능평가를 수행하는 것이 바람직하지만 비용, 시간, 대상이 사람이라는 점 등으로 인한 한계점이 있다. 많은 연구에서 어음처리에 따른 청력회복의 정도를 acoustic simulation에 기반하여 평가, 예측하고 있다. Acoustic simulation은 정상청각자를 대상으로 특정 어음처리방식에 기반하여 인공와우 시술자에게 들릴 것으로 예상되는 음성을 합성하여 들려주어 얻은 어음인지율을 토대로 어음처리방식을 평가하는 방법이다 [8,22].

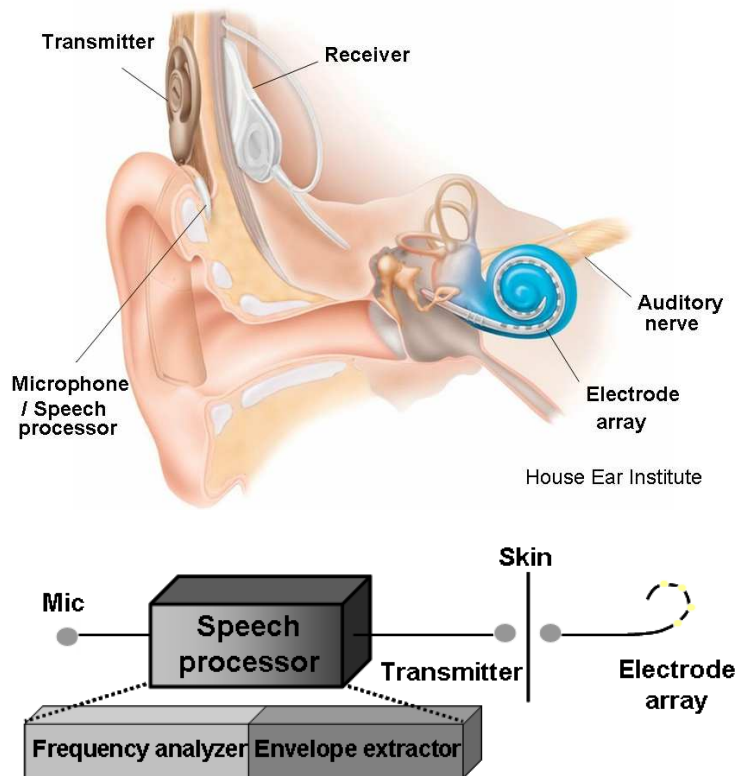


그림 1. 인공와우의 구조  
(a) 인공와우 이식도 (b) 인공와우 구조 블록도

Acoustic simulation은 내이에 전달되는 음성의 시간에 따른 정보 (temporal information)가 음성신호의 포락선 (envelope)과 spectral cue에 의해 결정된다는 가정 하에 수행된다. 그러나 음성에는 이 실험에서 고려한 포락선 성분 외의 것이 존재하며 이 실험을 통해서도 영아들의 경우 단일 채널 인공와우로도 청력회복이 유효하거나 일부 인공와우 수혜자가 음악을 들을 수 있을 정도의 청력회복을 보이는 경우 등에 대하여 만족할 만한 설명을 제시하는 것이 불가능하다 [8].

## 1.2 연구내용 및 범위

인공와우 어음처리방식의 성능 평가는 실제 이식과 재활훈련 후에 수행하는 것이 이상적이다. 그러나 비용, 시간 그리고 시술대상이 사람이라는 것 등의 제약점으로 인하여 모의실험을 통한 성능평가가 이루어지고 있다. 많은 연구에서 spectrogram 분석, acoustic simulation 등을 이용하여 성능평가를 수행하고 있으나 한계점이 존재하며 이를 극복하기 위하여 다양한 평가방법을 이용한 검증이 요구된다. 다양한 평가방법을 통하여 우수성이 검증된 어음처리방식이 실제 인공와우 이식 후에도 우수한 성능을 보일 여지가 많기 때문이다.

본 연구에서는 새로운 어음처리방식 평가방법으로 spike train decoding에 기반한 방법을 제안한다. 외부자극은 감각기 (receptor) 및 이에 연결된 감각신경에 의해 신경세포의 활동전위 (action potential) 형태로 변환된다. 활동전위의 발생시점은 spike train의 형태로 표현할 수 있는데 이로부터 외부자극에 대한 정보를 추출해내는 과정을 spike train decoding이라 한다 [9,10,14].

우선 일반적으로 사용되고 있는 시불변-선형 필터뱅크 기반 어음처리방식과 잡음하에서 강인한 음성정보 전달특성을 갖도록 어음처리방식의 주파수분석 부분이 수정된 시변-비선형 필터뱅크 기반 어음처리방식의 성능평가를 수행하였다. 시변-비선형 필터뱅크 기반 어음처리방식으로는 Meddis 등이 제안한 Dual Resonance Nonlinear Model (DRNL)을 채택하여 제작된 프로그램을 사용하였다 [15,16,22]. Deng과 Geisler는 spectrogram 분석 등을 통하여 비선형 특성을 가진 기저막 필터가 잡음 하에서 강인한 음성정보 전달특성을 가짐을 보였는데 DRNL

은 이러한 특성을 시변-비선형 필터뱅크로 표현한 모델이다 [26]. 그러므로 spike train decoding에 기반한 어음처리방식의 평가를 통하여 현재 널리 이용되는 어음처리방식에 비하여 시변-비선형 필터뱅크 기반 어음처리방식의 잡음 하 강인한 음성정보 전달특성을 보일 것으로 기대할 수 있다.

두 어음처리방식의 성능평가를 위하여 음성에 대하여 확률적 청신경모델로부터 multiple spike trains을 얻고 spike train decoding에 기반하여 외부자극인 입력음성의 포먼트 추정값을 얻는다. 이 때 특정 어음처리방식에 기반하여 생성된 전기자극펄스를 실제 사람의 청신경에 인가하여 얻은 spike train을 이용하는 것이 실제적인 성능평가를 위하여 이상적이다. 그러나 사람을 대상으로 실험하기에는 제약점이 있으므로 Bruce 등이 고양이의 청신경 응답특성을 모사한 청신경 모델을 이용하였다 [20]. 모델을 이용할 경우 실제 청신경을 대상으로 하는 실험의 경우와 달리 청신경의 역치나 불응기 함수 등의 특성을 조절할 수 있다는 장점이 있다. 청신경 모델로부터 얻은 spike train을 이용하여 추정한 포먼트를 원래의 포먼트와 정량적으로 비교함으로써 사용된 어음처리방식의 성능을 평가한다. 이는 입력음성의 특성이 지정된 어음처리방식에 기반한 전기자극펄스에 의해 청신경의 응답으로 잘 전달될수록 우수한 어음처리방식이라 할 수 있으며 높은 수준의 청력회복을 기대할 수 있다는 가정에 기반하고 있다.

음성을 인식하는데 있어서 모음은 중요한 매우 중요한 역할을 한다. 음성 주파수 성분의 정점값인 포먼트는 모음에 대한 중요한 단서를 제공한다. 포먼트의 낮은 값부터 제1, 제2 포먼트 (F1, F2)의 순서로 불리우는데 이 중 모음의 인식에는 제1, 제2 포먼트가 중요한 역할을 한다 [11,12]. 또한 Sachs 등에 의하면 음성의 포먼트 성분은 청신경의 응답으로 전달되며 청신경의 응답으로부터 포먼트를 추정해낼 수 있다. 그러므로 임의의 음성정보를 담은 전기자극펄스에 의하여 자극된 청신경의 응답으로부터 음성의 포먼트 성분을 추정할 수 있으며 본 연구에서 제안한 어음처리방식 성능평가의 지표로써 음성의 포먼트를 이용하는 것의 타당성을 뒷받침한다. 본 연구에서 시불변-선형 필터뱅크 기반 어음처리방식과 시변-비선형 필터뱅크 기반 어음처리방식의 성능평가 실험에는 제1~제3 포먼트 (F1, F2, F3)를 어음처리방식의 성능평가의 지표로 사용하였다.



다음으로 어음처리방식에 적응효과 (adaptation)를 추가하여 입력음성의 포락선 추출특성을 개선할 경우 음성의 시작점에 대한 정보 전달이 효과적으로 이루어지는지 관찰하였다. 적응효과는 내측유모세포 (Inner hair cell; IHC)와 청신경 사이의 화학적 시냅스 (chemical synapse)에 기원하는 현상으로 자극의 시작점에서 신경전달물질 (neurotransmitter)의 급격한 방출로 높은 발화율을 보이고 이후 신경전달물질의 고갈 (depletion)에 의하여 발화율이 단조 감소하는 형태를 보인다 [3,23,24]. 정상인의 경우 적응효과가 발생하지만 청신경을 전기적으로 자극하여 소리정보를 전달하는 인공와우의 경우 이를 기대할 수 없다 [22-24]. 어음처리방식에 적응효과를 적용할 경우 음성시작점의 정보 전달특성을 spike train decoding에 기반한 방법을 이용하여 평가하였다. 적응효과를 적용하기 전후의 어음처리방식에 의하여 생성된 전기자극펄스를 청신경 모델에 인가하여 spike train을 얻는다. 이로부터 계산된 발화율 행렬과 음성의 유무를 표시한 행렬을 이용하여 음성유무를 판별하는 선형분류기를 얻는다 [25]. 새로이 얻은 발화율 행렬과 선형분류기를 이용하여 음성의 시작점에서 음성유무를 분류한다. 음성의 시작점에 대하여 음성유무 분류를 수행하였으므로 음성이 있다고 판별하는 경우가 많을수록 음성시작점 정보 전달특성이 우수하다고 볼 수 있다. 이와 같은 방법을 이용하여 어음처리방식에 적응효과를 추가함으로써 음성의 시작점에 대한 정보를 효과적으로 전달할 수 있는지 알아보았다.

## 제 2 장 방 법

### 2.1 Spike train decoding

외부자극은 감각기 (receptor) 및 이에 연결된 감각신경에 의해 신경세포의 활동전위 (action potential) 형태로 변환된다. 활동전위의 발생시점은 spike train의 형태로 표현할 수 있는데 이로부터 외부자극에 대한 정보를 추출해 내는 과정을 spike train decoding이라 한다 [9,10,14].

그림 2.2는 최적선형 디코딩필터를 이용한 spike train decoding 과정을 나타낸 그림이다. 디코딩 필터는 spike train들로부터 일정 time bin 내의 spike 발생 개수를 계산한 발화율 (firing rate)을 입력으로 하고 외부자극을 출력으로 한다. 디코딩 필터의 입출력 간의 매핑 즉, 디코딩 필터계수를 결정하는 과정을 학습이라고 하는데 학습과정은 주어진 외부자극의 참값과 추정값 간의 평균 제곱 오차 (mean squared error)를 최소화하도록 이루어지며 학습으로부터 얻어진 디코딩 필터를 이용하여 새로운 spike train으로부터 외부 자극을 추정한다. 디코딩 필터는 다중 채널 유한임펄스 응답 필터 (multichannel finite impulse response filter) 형태로 구현되어 있으며 다음과 같이 입력자극과 발화율 정보를 기반으로 하여 필터 계수들을 구할 수 있다.

$$\mathbf{f} = [f_1(0) f_1(1) \cdots f_1(M-1) \cdots f_p(0) \cdots f_p(M-1)]^T \quad (2.1)$$

$$\mathbf{f} = (\mathbf{R}^T \mathbf{R})^{-1} \cdot (\mathbf{R}^T \mathbf{s}) \quad (2.2)$$

여기서  $f_p(j)$ 는 뉴런  $p$ 의  $j$ 번째 time bin에 대한 필터계수를 의미한다.  $\mathbf{R}$  및  $\mathbf{s}$ 는 각각 발화율 및 입력자극에 대한 정보를 포함하고 있는 행렬로 다음과 같이 구성된다.

$$\mathbf{R} = \begin{bmatrix} 1 & r_1(0) & r_1(1) & \cdots & r_1(M-1) & r_N(0) & \cdots & r_N(M-1) \\ 1 & r_1(1) & r_1(2) & \cdots & r_1(M) & r_N(1) & \cdots & r_N(M) \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & r_1(L) & r_1(L+1) & \cdots & r_1(L+M-1) & r_N(L) & \cdots & r_N(L-M+1) \end{bmatrix} \quad (2.3)$$

$$\mathbf{s} = [s(0) s(1) \cdots s(L-1)]^T \quad (2.4)$$

$r_N(i)$ 는  $N$ 번째 단일 뉴런이  $i$ 번째 time bin에서 갖는 평균발화율을 뜻하며  $M$ 은 디코딩 필터의 차수,  $N$ 은 관찰된 단일 뉴런의 개수 그리고  $L$ 은 총 관찰시간 내 time bin의 개수이다. 학습 과정에서 얻은 디코딩 필터계수와 새로운 spike train을 이용하여 spike train에 인코딩된 외부 자극을 복원하는 과정을 추정이라 한다. 외부자극의 추정과정은 다음과 같이 표현할 수 있다.

$$\hat{s}(i) = \sum_{p=1}^N \sum_{j=0}^{M-1} r_p(i-j) f_p(j) \quad (2.5)$$

변수  $r_p(i)$ 는 뉴런  $p$ 의  $i$ 번째 time bin에 대한 발화율,  $\hat{s}(i)$ 는  $i$ 번째 time bin에 대한 외부자극의 추정값을 의미한다.

## 2.2 어음처리방식 성능평가방법

실험에 이용하기 위하여 시불변-선형 필터뱅크 및 시변-비선형 필터뱅크 기반 어음처리방식 그리고 각 어음처리방식에 적응효과를 더할 경우에 대하여 전기자극펄스를 생성하였다. 어음처리방식을 통한 전기자극펄스 생성을 위하여 선행 연구를 통하여 제작된 프로그램을 이용하였다 [22].

### 2.2.1 포먼트 디코딩을 통한 어음처리방식 성능평가

그림 2.1은 spike train decoding에 기반한 어음처리방식 평가방법을 설명하기 위한 그림이다. 확률적 청신경모델 [18-20]에 특정 어음처리방식에 기반한 전기자극펄스를 인가하여 여러 개의 청신경으로부터 spike train들을 얻는다. 그림 2.2는 최적선형필터를 이용한 spike train decoding을 설명한 그림이다. Spike train으로부터 발화율을 계산하여 최적선형 디코딩필터의 입력으로 하고 입력음성으로부터 얻은 제1~제3 포먼트 궤적 (F1~F3)의 참값을 출력으로 사용하여 디코딩 필터계수를 결정한다. 얻어진 디코딩필터의 입력으로 입력음성에 대한 청신경모델의 응답

을 사용하여 포먼트의 추정값을 얻고 이를 참값과 상관계수를 이용하여 비교하여 포먼트가 얼마나 충실하게 복원되는지를 정량적으로 관찰한다. 포먼트 궤적이 충실하게 복원될수록 참값과 유사한 형태를 가지며 상관계수가 1에 가까운 값을 갖게 되며 이를 이용하여 디코딩 성능을 정량화하였다. 학습을 통해 얻어진 디코딩 필터의 계수와 청신경모델 응답의 통계적 특성 등에 대하여 독립적인 성능을 관찰하기 위하여 50회의 시뮬레이션을 통해 얻은 상관계수의 평균값을 비교, 관찰하였다. 평가할 어음처리방식으로 시블변-선형 필터뱅크 기반 어음처리방식과 시변-비선형 필터뱅크 기반 어음처리방식을 이용하였고, 여러 가지 잡음 하에서 각 어음처리방식에 대한 포먼트 디코딩 성능을 비교하고 paired t-test를 통하여 성능의 차이가 통계적으로 유의한가를 알아보았다.

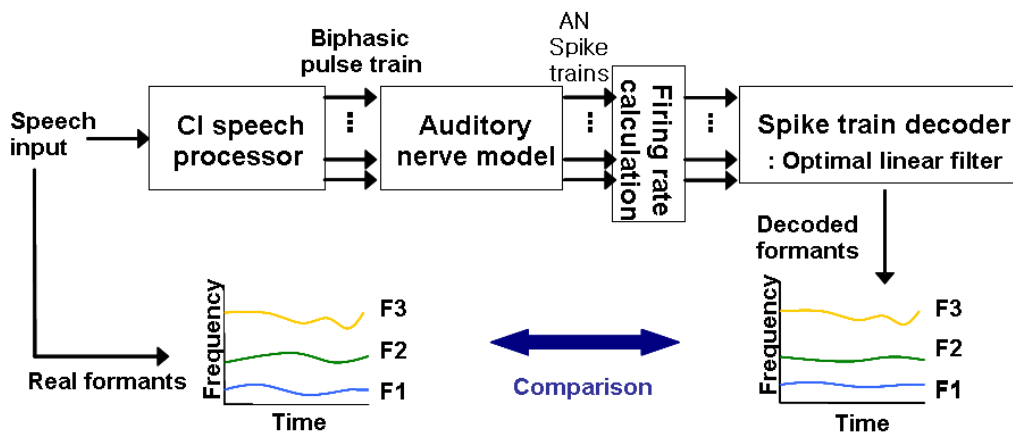


그림 2.1 Spike train decoding 기반 포먼트 디코딩방법

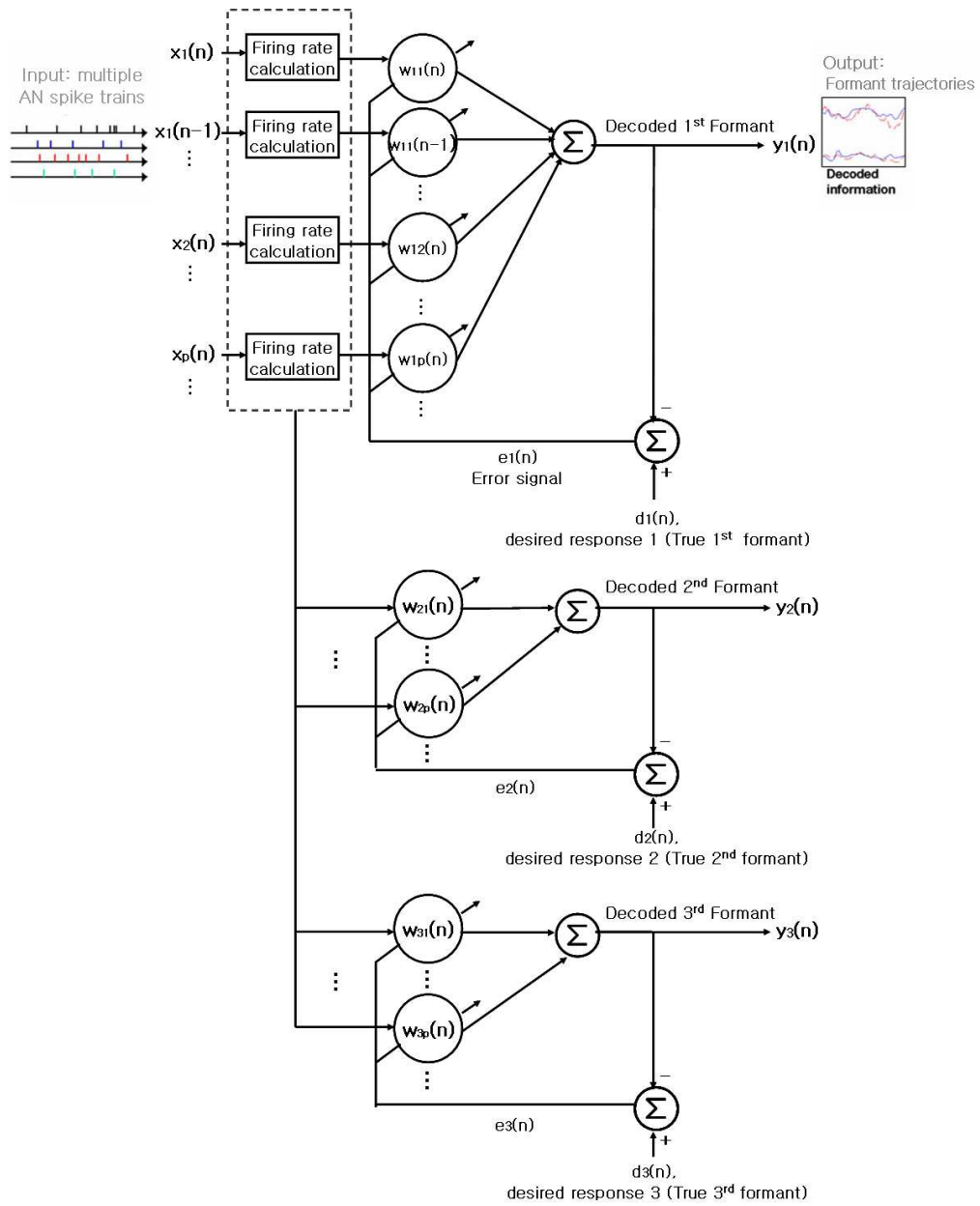


그림 2.2 최적선형필터를 이용한 spike train decoding

## 2.2.2 음성시작점 정보 전달특성 평가

그림 2.3은 spike train decoding에 기반하여 음성시작점의 정보 전달특성을 평가하는 방법을 설명하기 위한 그림이다. 적응효과를 적용 전후의 어음처리방식 [22-24]에 기반한 전기자극펄스에 의하여 자극된 청신경 모델 응답과 음성의 유무를 각각 class 1, class -1로 할당한 행렬을 이용하여 Wiener solution에 기반한 최적 선형분류기를 얻는다 [25]. 선형분류기의 입력으로 청신경 모델 응답을 이용하여 음성의 시작점에 대한 음성유무 분류결과를 얻고 이를 참값과 비교하여 적응효과 적용 전후의 음성시작점 정보 전달특성을 비교한다. 음성유무 행렬의 참값은 그림 3에 점선으로 표시된 바와 같이 음성이 있는 부분은 1, 그렇지 않은 부분은 -1로 표시하였고, 추정값은 점실선으로 표시된 바와 같다. 음성시작점은 음성 제작시 실제 음성이 시작되는 부분부터 20 ms동안으로 가정하였다 [23].

포먼트 디코딩의 경우 디코딩필터로 최적선형필터를 이용하였고 디코딩필터의 출력으로 포먼트 궤적을 이용하였다. 반면 음성시작점에서 음성유무 분류를 통한 음성시작점에 대한 정보의 전달특성 평가를 위해서는 선형분류기를 디코딩필터로 이용하였고 음성유무를 나타낸 행렬을 출력으로 이용하였다. 그림 2.4는 선형분류기를 이용하여 음성시작점에서 음성유무 분류를 수행하기 위한 spike train decoding을 설명한 그림이다. 청신경모델 응답으로부터 계산된 발화율 행렬을 디코딩필터의 입력으로 이용하고 음성이 있는 부분은 1, 없는 부분을 -1로 표시한 행렬을 출력으로 이용하여 Wiener solution에 기반한 최적선형분류기를 얻는다. 선형분류기는 식 (2.6)을 이용하여 얻을 수 있다.

$$W=R^{-1}P, P=E[dR], R=E[XX^T] \quad (2.6)$$

이 과정으로부터 얻은 선형분류기와 새로운 발화율 행렬  $X_2$ 를 이용하여 식 (2.7)에 표기한 바와 같이 음성유무 분류행렬의 추정값을 얻을 수 있고, 음성시작점에 해당하는 부분만 음성유무 성능평가에 이용한다.

$$\hat{d}=X_2W, \begin{cases} \hat{d}> 0, class 1 \\ \hat{d}< 0, class -1 \end{cases} \quad (2.7)$$

발화율 행렬은  $X$ , 음성유무를 표기한 행렬  $d$ 는 식 (2.8)과 같이 표현할 수 있다.

$N$ 은 관찰된 단일 뉴런의 개수,  $r_N(i)$ 는  $N$ 번째 단일 뉴런이  $i$ 번째 time bin에서 갖는 평균발화율을 뜻하며  $L$ 은 총 관찰시간 내 time bin의 개수이다.

$$X = \begin{bmatrix} 1 & r_1(0) & r_2(0) & \dots & r_N(0) \\ 1 & r_1(1) & r_2(1) & \dots & r_N(1) \\ \dots & \dots & \dots & \dots & \dots \\ 1 & r_1(L) & r_2(L) & \dots & r_N(L) \end{bmatrix} \quad (2.8)$$

$$d = [d(0) d(1) d(2) \dots d(L-1)] \quad (2.9)$$

어음처리방식으로 시불변-선형 필터뱅크 기반 어음처리방식과 시변-비선형 필터뱅크 기반 어음처리방식을 이용하였으며 각 어음처리방식에 적응효과 적용 전후의 음성시작점 정보 전달특성을 평가하였다. 20회의 시뮬레이션을 통하여 음성시작점 정보 전달특성 평가에 이용된 음성유무분류 샘플 수는 160개 이며 이 중 음성유무를 올바르게 분류한 비율을 통하여 평가하였다. 음성시작점 정보 전달특성을 정량적으로 평가하기 위한 '% correct'는 다음과 같이 계산되었다.

$$\% correct = \frac{\text{음성을 올바르게 분류한 경우의 개수}}{160} (\%) \quad (2.10)$$

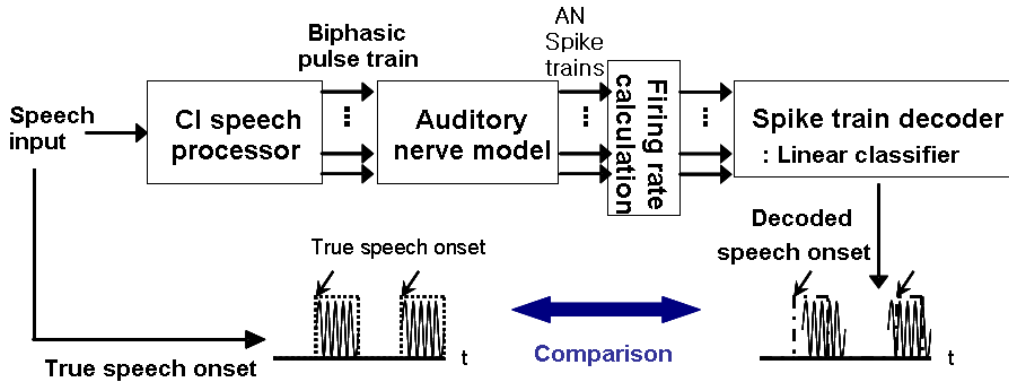


그림 2.3 Spike train decoding 기반 음성시작점 정보 전달특성 평가방법

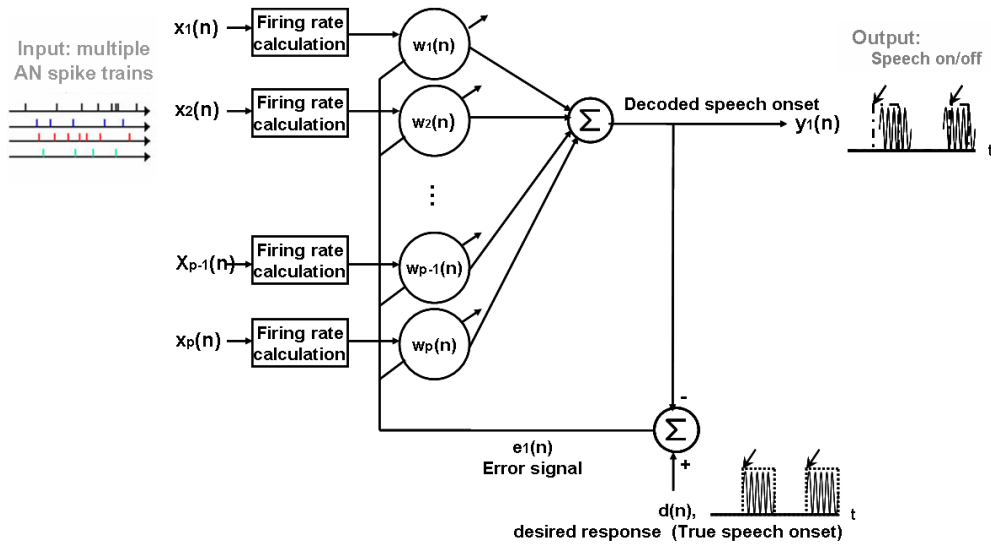


그림 2.4 선형분류기를 이용한 spike train decoding

## 2.3 어음처리방식의 종류

### 2.3.1 시불변-선형 필터뱅크 기반 어음처리방식

시불변-선형 필터뱅크 기반 어음처리방식은 현재 가장 널리 사용되고 있으며 달팽이관 내 기저막 (basilar membrane)의 특성을 선형 대역통과필터로 표현하고 있다. 잡음이 없는 경우 높은 어음 인식율을 보이나 잡음 하에서 사용자의 어음 인식율이 급격히 감소한다 [15-17,21,22].

### 2.3.2 시변-비선형 필터뱅크 기반 어음처리방식

시변-비선형 필터뱅크 기반 어음처리방식은 기저막의 비선형 특성을 모델링하기 위한 dual resonance nonlinear (DRNL) 모델에 기반하고 있으며 한 채널이 그림



2.5(c) 에서와 같이 선형, 비선형 두 개의 신호처리 통로로 구성되어 채널의 출력은 두 통로의 합이 된다 [16,21,22]. 선형 통로의 중심주파수는 비선형 통로의 중심주파수와 약간 어긋나 있어 쌍공명 (dual resonance)필터라 한다. 또한 대역 통과 필터의 대역폭이 비대칭적이고 compression과 같은 외측유모세포의 비선형성을 표현한다. 이와 같은 기저막의 비선형성에 의해 시변-비선형 필터뱅크 기반 어음 처리방식은 잡음 하에서도 음성 정보 전달 측면에서 강인한 특성을 보이는 것으로 알려져 있다 [16,17,21,22].

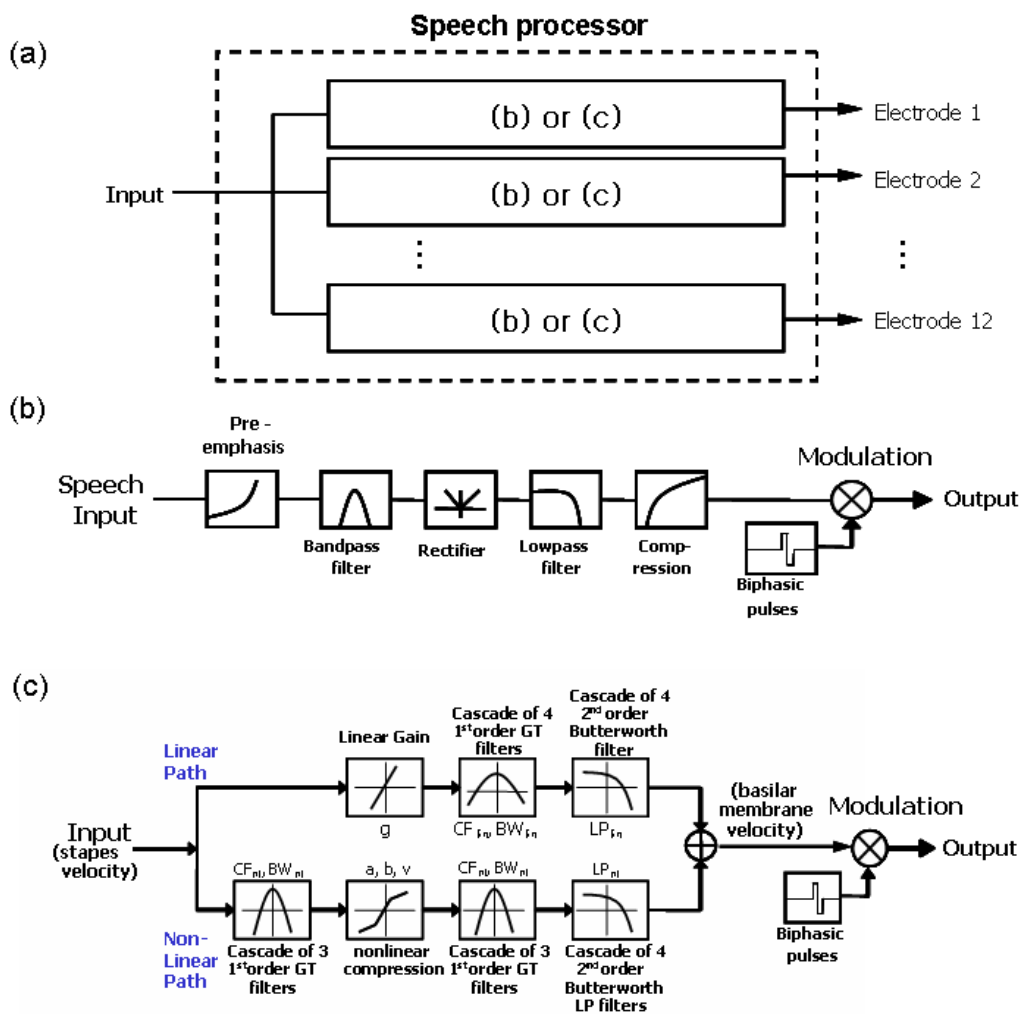


그림 2.5 어음처리방식에 기반한 전기자극펄스생성 (a) 어음처리방식((b) 또는 (c))에 기반한 채널별 전기자극펄스 생성 과정 블록도, (b) 시불변-선형 필터뱅크 기반 어음처리방식, (c) 시변-비선형 필터뱅크 기반 어음처리방식

### 2.3.3 적응효과

그림 2.6(a)에서 설명하는바와 같이 귀에 일정 크기의 자극 (소리)이 인가되었을 때 청신경의 응답은 자극의 시작점에서는 높은 발화율을 보이고 시간이 지남에 따라 감소하는 형태를 보인다. 이러한 청신경의 응답특성을 적응효과라 하며 달팽이관 내의 내측유모세포 (Inner hair cell; IHC)와 청신경 (Auditory nerve; AN) 사이의 화학적 시냅스에 기원한다. 자극 초기에 신경전달물질이 급격하게 방출되어 높은 발화율을 보이고 이후로 신경전달물질의 소모로 인하여 발화율은 단조감소하게 된다 [1-3]. 일반적인 인공와우 어음처리방식을 채택한 인공와우 이용자의 경우 청신경에 직접 전기자극을 가하므로 IHC/AN 사이의 화학적 시냅스에서 보이는 적응효과를 기대할 수 없다 (그림 2.6(b)). 그러므로 일반적으로 이용되고 있는 어음처리방식에 적응효과를 더하여 주어야 한다. 어음처리방식에 적용할 적응효과 알고리즘은 입력음성에 대하여 음성시작점을 검출한다. 그리고 시작점 부분에 해당하는 포락선 (envelope)에 이득값을 곱하여 원 포락선에 더해줌으로써 음성시작점을 강조해 준다 [22,23].

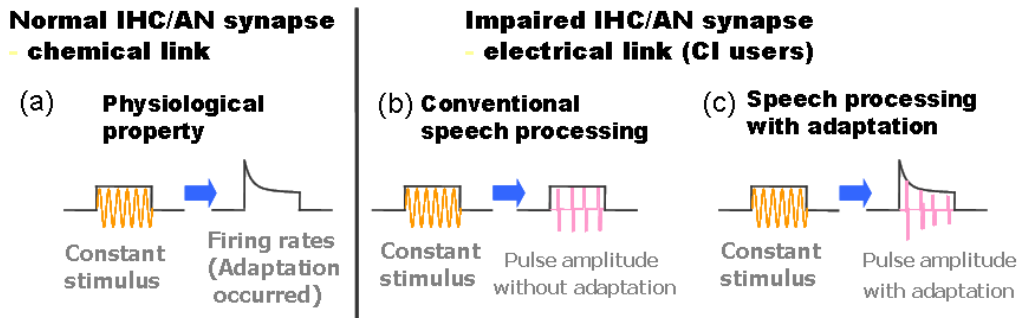


그림 2.6. IHC/AN 간 화학적 시냅스에 기원하는 적응효과. (a) 정상인의 IHC/AN 간 화학적 시냅스에 기원하는 적응효과, (b) 일반적으로 이용되는 어음처리방식에 기반하여 생성된 전기자극펄스, (c) 어음처리방식에 적응효과를 적용하여 생성된 전기자극펄스

## 2.4 확률적 청신경 모델

Spike train decoding에 기반하여 어음처리방식을 평가하기 위해 어음처리방식을 통하여 생성된 전기자극펄스를 청신경 모델에 인가하여 spike train을 얻는 과정이 필요하다. 어음처리방식을 통하여 생성된 전기자극펄스를 실제 사람의 청신경에 인가하여 얻은 spike train을 이용하는 것이 실제적인 성능평가를 위하여 이상적이거나 사람을 대상으로 실험하기에는 제약점이 있으므로 Bruce 등이 고양이의 청신경 응답특성을 모사한 청신경 모델을 이용하였다 [20]. 또한 실제 청신경을 대상으로 하는 실험의 경우와 달리 모델을 이용할 경우 청신경의 역치나 불응기 함수 등의 특성을 조절할 수 있다는 장점이 있다. 확률적 청신경모델 (그림 2.7)은 특정 어음처리방식에 의해 생성된 전기자극펄스를 입력으로 하며 spike train을 출력으로 한다 [18,19,20]. 전기자극펄스에 실린 소리의 정보가 청신경의 spike train으로 전달되므로 소리의 정보가 청신경 응답에 어떻게 인코딩 되는지를 관찰할 수 있다. 모델의 응답을 이용한 spike train decoding을 수행함으로써 어음처리방식의 성능평가를 수행할 수 있다. 확률적 청신경모델은 전기자극펄스의 세기가 청신경

의 세포막 노이즈와 불응기를 고려한 문턱치의 합보다 클 경우 발화하도록 되어 있다. 그림 2.7에서  $V_{stim}$ 은 전기자극펄스 입력,  $V_{noise}$ 는 2500 Hz 대역 제한된 백색 잡음이며 이것의 표준편차는 modulation depth와 같은 비율로 결정하였다 (modulation depth/ $V_{noise}$ 의 표준편차 = 1).  $V_{th}$ 은 청신경의 문턱치,  $V_{ref}$ 는 불응기를 고려한 청신경의 문턱치 변화이다.

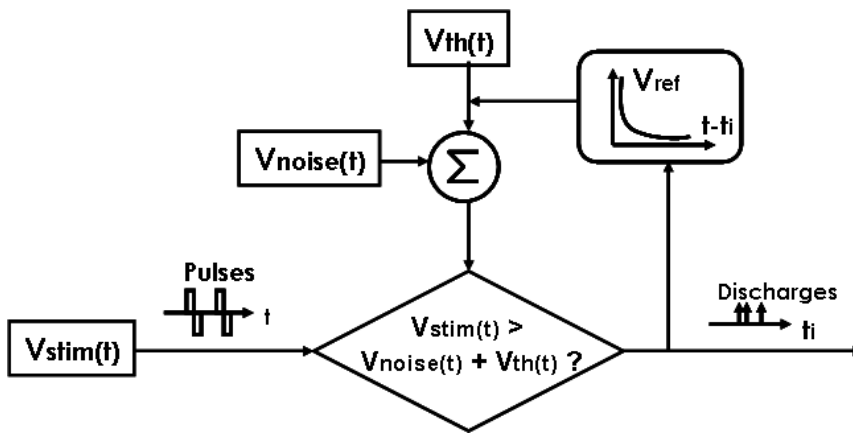


그림 2.7 확률적 청신경모델.  $V_{stim}(t)$ : 특정 어음처리방식에 기반하여 생성된 전기 자극펄스열,  $V_{noise}(t)$ : 2500Hz 대역 제한된 백색잡음,  $V_{th}(t)$ : 불응기를 고려한 청신경모델의 문턱치

## 2.5 시뮬레이션을 위한 음성제작방법

### 2.5.1 포먼트 디코딩을 위한 음성제작

시뮬레이션시 ‘자음+모음’ 형태의 합성음 /mu/가 6회 반복 표현된 1.7 초의 길이의 음성을 제작하였다. 잡음 하의 경우 음성 파형에 대하여 5 dB SNR, 0 dB SNR의 크기를 갖는 백색잡음 (white Gaussian noise; WGN)과 음성형태 (speech shaped noise; SSN)을 첨가하였다. 백색잡음은 모든 대역의 주파수 성분을 갖는 잡음이고, 음성형태 잡음은 백색잡음에 대하여 1100 Hz의 cutoff 주파수를 갖는 저역대역통과필터를 통과시켜 음성과 유사한 형태의 스펙트럼을 갖도록 제작한 잡음이다 [27].

### 2.5.2 음성시작점 정보 전달특성 평가를 위한 음성 제작

실험에 이용된 음성은 ‘자음+모음’ 형태이고 무음과 음성이 200 ms씩 반복되며 800 ms 길이를 가진다. 자음은 유성파열음 (voiced stop) /b/, /d/와 무성파열음 (unvoiced stop) /p/, /t/ 그리고 유성마찰음 (voiced fricative) /v/, /z/와 무성마찰음 (unvoiced fricative) /f/, /s/ 를 이용하였다. 모음은 /a/를 이용하여 /ba/, /da/, /pa/, /ta/, /va/, /za/, /fa/, /sa/ 와 같은 음성을 제작하여 실험에 사용하였다. 잡음 하 어음처리방식 성능평가를 위한 백색잡음과 음성형태잡음은 포먼트 디코딩의 경우와 같은 방법으로 제작되었다.

## 제 3 장 결 과

### 3.1 포먼트 디코딩을 통한 어음처리방식 성능평가

#### 3.1.1 포먼트 디코딩에 이용된 상세 파라미터

음성을 사용하여 그림 2.5(a)에 나타낸바와 같이 12채널에 대한 전기자극펄스를 생성하였다. 이 때 전기자극펄스의 자극율은 800, 2000, 5000 pulse per second (pps)로 설정하였고 펄스폭은 각각 100, 100, 25  $\mu\text{s}/\text{phase}$ 로 설정하였다. 진폭 변조 (amplitude modulation)는 최소값을 441  $\mu\text{A}$ 로 두고 1% modulation depth를 갖도록 조절하였다. Spike train decoding을 수행하기 위해 디코딩필터로 최적선형필터를 이용하였으며 12채널로부터 얻은 spike train의 발화율 계산을 위한 time bin 폭과 디코딩필터의 차수는 가장 높은 디코딩 성능을 보이되 학습 데이터에 대하여 overfitting되지 않도록 최적화 되었다. Time bin 폭은 15 ms로 설정하였고 디코딩필터의 차수는 5차로 75 ms 길이를 갖도록 설정하였다. 실험에 사용된 잡음 환경은 잡음이 없는 경우, 5 혹은 0 dB SNR 백색잡음 (white Gaussian noise; WGN) 하 그리고 5 dB SNR 음성형태 잡음 (음성과 유사한 형태의 스펙트럼을 갖는 잡음 - speech shaped noise; SSN) 하 등으로 각각의 조건에서 두 어음처리방식의 성능평가 비교를 수행하였다.

#### 3.1.2 포먼트 궤적의 참값과 추정값 간 비교

그림 3.1은 전기자극펄스의 자극율이 2000인 경우 입력음성으로부터 추출된 제1~제3 포먼트 궤적 (F1~F3)의 참값과 추정값을 비교한 결과이다. 여러 번의 시뮬레이션으로부터 얻은 그림 중 보편적으로 보이는 결과를 선택하였다. 잡음이 없는 경우 혹은 5 dB SNR 백색잡음 (white Gaussian noise, WGN) 하에서 두 자극방

법 간 차이는 거의 없는 것으로 관찰된다. 그러나 잡음레벨이 상대적으로 큰 0 dB SNR 백색잡음 하 및 입력음성 형태의 잡음(speech shaped noise, SSN) 하의 경우 시변-비선형 필터뱅크 기반 어음처리 방식에 의한 포먼트 궤적의 추정값이 참값과 더 유사해 보인다. 포먼트 궤적의 참값과 추정값의 차이를 육안상으로 관찰하여 어음처리방식의 성능평가를 수행하기는 쉽지 않으므로 유사성을 정량적으로 평가하는 것이 필요하다. 다음 절에서는 포먼트 궤적의 참값과 추정값 간의 유사성을 상관계수로 정량화하여 나타낸 결과를 보인다.

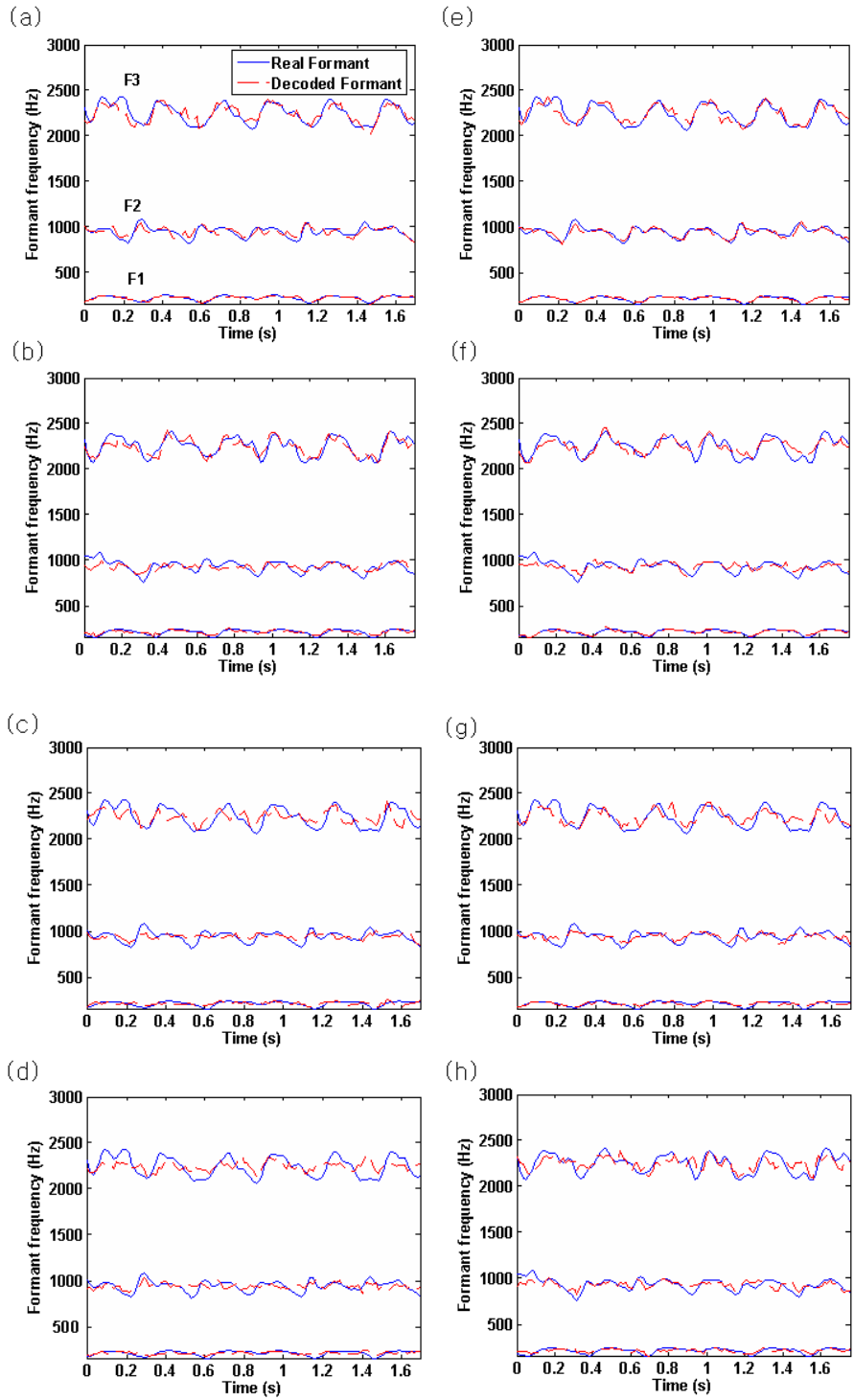




그림 3.1 포먼트 궤적의 참값 (실선)과 추정값 (점선) 비교. (a)~(d)는 시불변-선형 필터뱅크 기반 어음처리방식에 대한 포먼트 궤적. (e)~(h)는 시변-비선형 필터뱅크 기반 어음처리방식에 대한 포먼트 궤적. (a),(e)는 잡음이 없는 경우, (b),(f)는 5 dB SNR WGN 하, (c),(g)는 0 dB SNR WGN 하, (d),(h)는 5 dB SNR SSN 하

### 3.1.3 어음처리방식 간 평균 포먼트 디코딩 성능비교

그림 3.2는 전기자극펄스의 자극율이 800 pps 일 때, 50회의 시뮬레이션 결과로부터 얻은 제1~제3 포먼트 궤적 (F1~F3)의 추정값과 참값 간의 유사성을 상관계수의 평균값을 비교한 결과이다. 잡음이 없는 경우와 5 dB SNR 백색잡음 하의 경우 성능의 평균값에 큰 차이가 나지 않으나 잡음레벨을 증가시켜 0 dB SNR 백색잡음 하 혹은 입력음성 형태의 잡음 하의 경우 시변-비선형 필터뱅크 기반 어음처리 방식의 우수성이 두드러졌다. 즉, 백색잡음에 대한 경우보다 음성 형태의 잡음 하 혹은 잡음 레벨이 증가할수록 시변-비선형 필터뱅크 기반 어음처리 방식의 우수성을 관찰되었다.

자극율이 2000 pps 일 때 (그림 3.3) 잡음이 없는 경우는 시불변-선형 필터뱅크 기반 어음처리방식과 시변-비선형 필터뱅크 기반 어음처리방식간의 성능의 평균값 차가 적었고, 5 dB SNR 백색잡음 하에서도 유사한 수준이었다. 그러나 백색잡음레벨이 0 dB SNR로 증가하거나 5 dB SNR 음성형태 잡음에서는 시변-비선형 필터뱅크에 의하여 포먼트가 더 충실히 복원되었으며 그 차이가 두드러졌다.

자극율이 5000 pps 일 때 (그림 3.4) 평균 포먼트 디코딩 성능은 다른 자극율에서와 마찬가지로 잡음이 없거나 5 dB SNR 백색잡음 하에서는 어음처리방식 간의 성능의 평균값 차가 크지 않았고, 잡음레벨이 상대적으로 큰 0 dB SNR 백색잡음 하에서 시변-비선형 필터뱅크 기반 어음처리방식의 우수성이 두드러졌다.

자극율이 5000일 경우 5 dB SNR 음성형태 잡음 하 F3를 제외한 대부분의 경우에서 두 어음처리방식 간의 성능차는 통계적으로 유의하였다 ( $p < 0.001$ ).

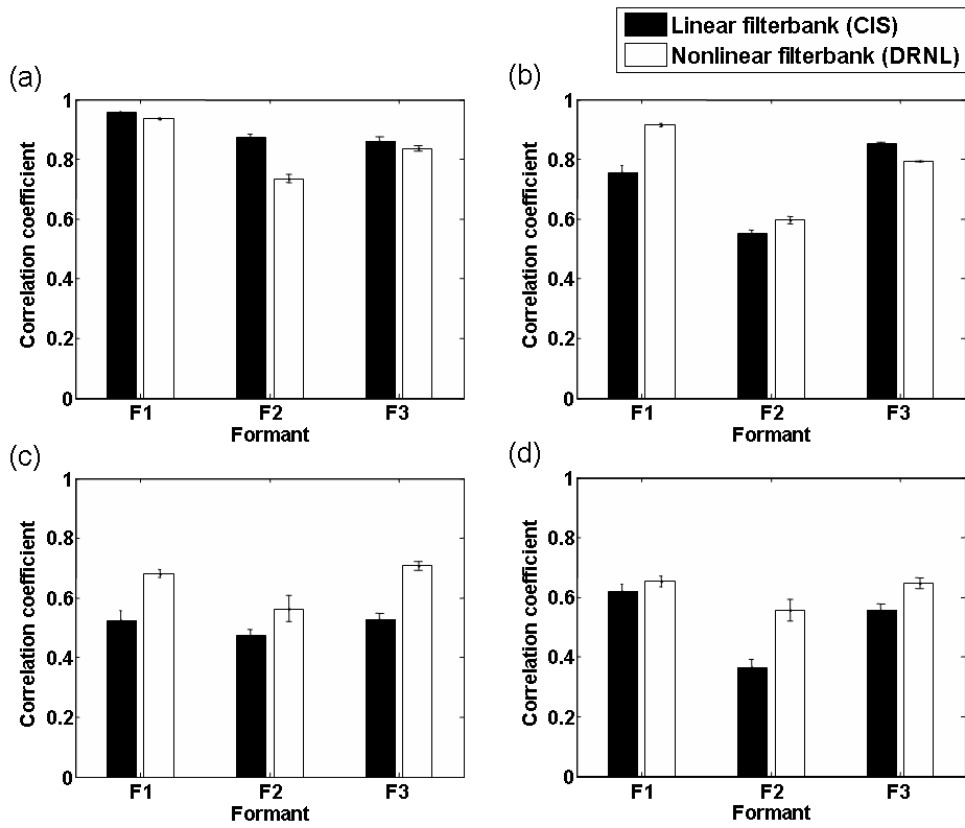


그림 3.2 어음처리방식 간 평균 포먼트 디코딩 성능비교 (자극율이 800 pps일 경우). (a)는 잡음이 없는 경우, (b)는 5 dB SNR 백색잡음 하, (c)는 0 dB SNR 백색잡음 하, (d)는 5 dB SNR 음성형태 잡음 하

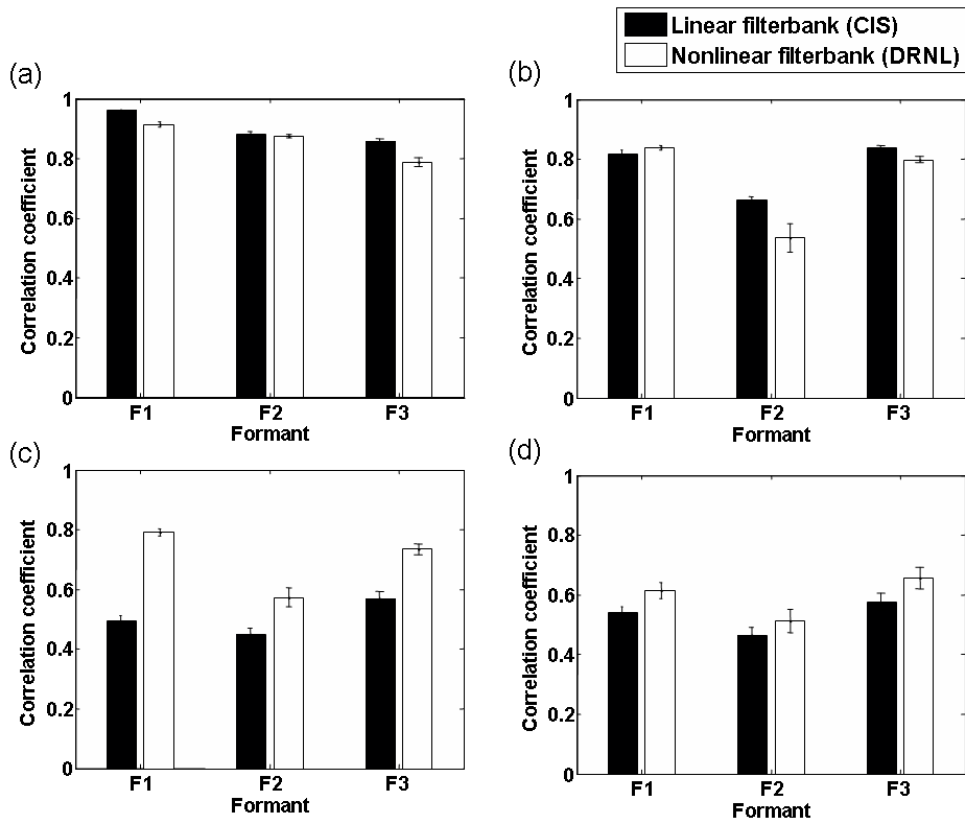


그림 3.3 어음처리방식 간 평균 포먼트 디코딩 성능비교 (자극율이 2000 pps일 경우). (a)는 잡음이 없는 경우, (b)는 5 dB SNR 백색잡음 하, (c)는 0 dB SNR 백색잡음 하, (d)는 5 dB SNR 음성형태 잡음 하

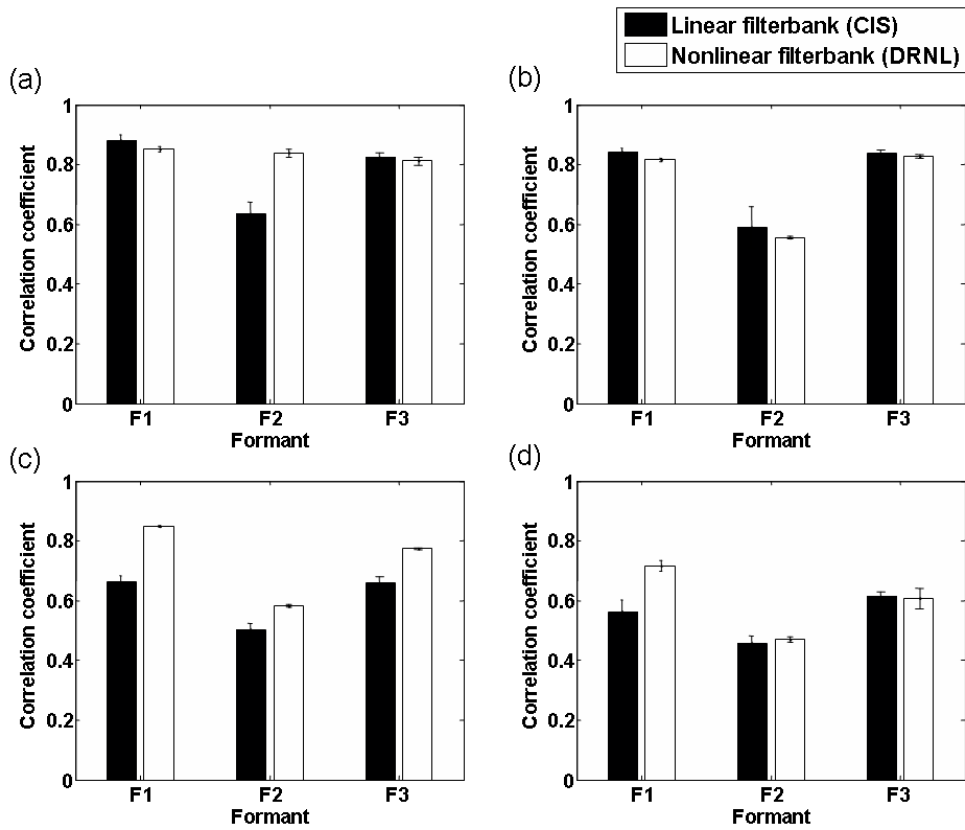


그림 3.4 어음처리방식 간 평균 포먼트 디코딩 성능비교 (자극율이 5000 pps일 경우). (a)는 잡음이 없는 경우, (b)는 5 dB SNR 백색잡음 하, (c)는 0 dB SNR 백색잡음 하, (d)는 5 dB SNR 음성형태 잡음 하

### 3.1.4 여러 가지 잡음 하에서 평균 포먼트 디코딩 성능비교

그림 3.5는 전기자극펄스의 자극율이 800 pps 일 때 여러 가지 종류의 잡음 하에서 포먼트 궤적의 추정값과 참값 간 유사성을 상관계수를 이용하여 정량적으로 비교한 결과이다. 시불변-선형 필터뱅크 기반 어음처리방식을 사용했을 경우 높은 SNR의 백색잡음 하와 입력음성형태의 잡음 하에서의 포먼트 디코딩 성능이 급격하게 감소하였다. 반면 시변-비선형 필터뱅크 기반 어음처리방식은 잡음에 대하여 강인한 특성을 보였다. 잡음이 없는 경우와 5 dB SNR 백색잡음 하에서는 시변-비선형 필터뱅크 기반 어음처리 방식과 유사한 수준의 평균성능을 보였으나 백색잡음레벨이 5 dB SNR에서 0 dB SNR로 증가한 부분과 음성형태의 잡음 하의 평균 디코딩 성능을 비교해보면 시변-비선형 필터뱅크 기반 어음처리방식에 의한 평균 디코딩 성능의 감소가 상대적으로 완만하다. 이와 같은 현상은 자극율이 2000 pps 일 때의 그림 3.6에서도 관찰된다. 자극율이 5000 pps 일 때 그림 3.7(a)에서는 시불변-선형 필터뱅크 기반 어음처리방식을 이용할 경우 잡음레벨 증가에 따른 포먼트 디코딩 성능이 800 혹은 2000 pps 의 경우보다는 완만하게 감소하지만 다른 자극율에서 보인바와 마찬가지로 시변-비선형 필터뱅크 기반 어음처리방식을 이용할 경우 포먼트 디코딩 성능이 개선됨을 볼 수 있다.

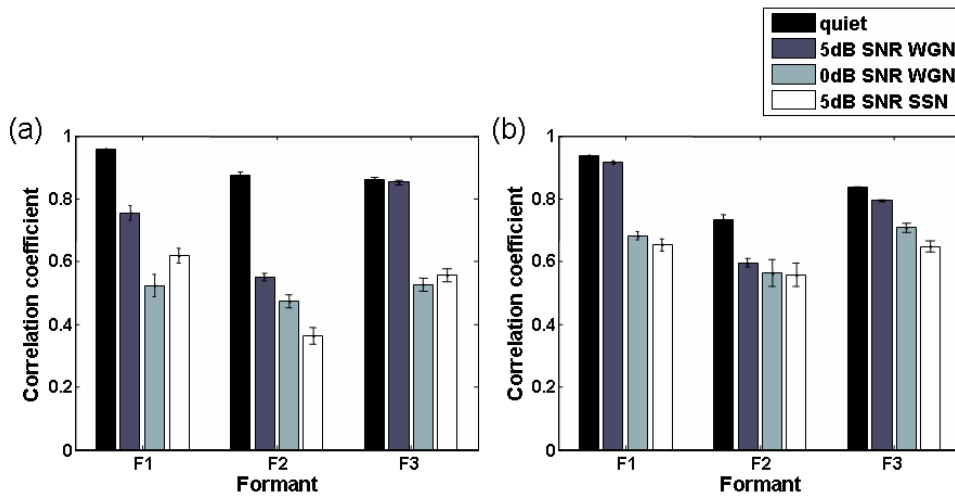


그림 3.5 여러 가지 종류 잡음 하의 평균 포먼트 디코딩 성능변화 (자극율이 800 pps일 경우). (a)는 시블변-선형 필터뱅크 기반 어음처리방식, (b)는 시변-비선형 필터뱅크 기반 어음처리방식에 대한 평균 포먼트 디코딩 성능

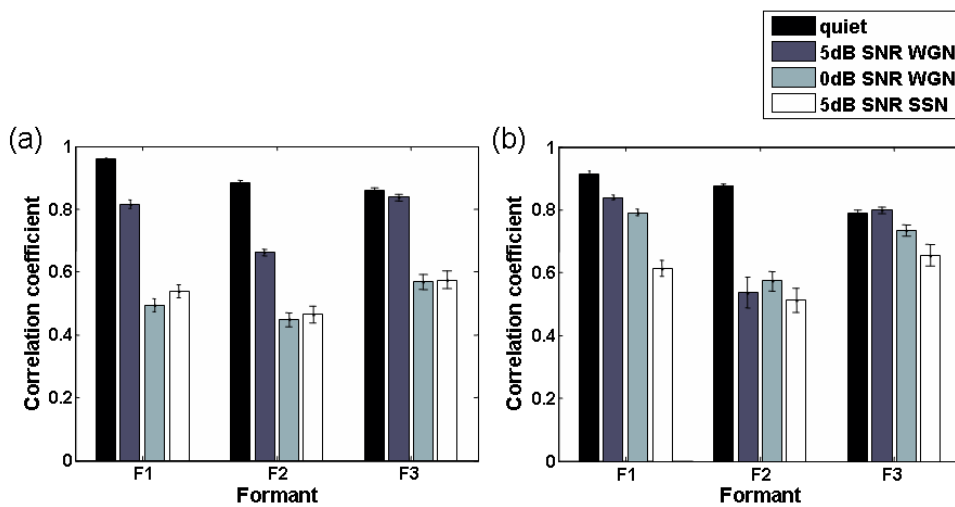


그림 3.6 여러 가지 종류 잡음 하의 평균 포먼트 디코딩 성능변화 (자극율이 2000 pps일 경우). (a)는 시블변-선형 필터뱅크 기반 어음처리방식, (b)는 시변-비선형 필터뱅크 기반 어음처리방식에 대한 평균 포먼트 디코딩 성능

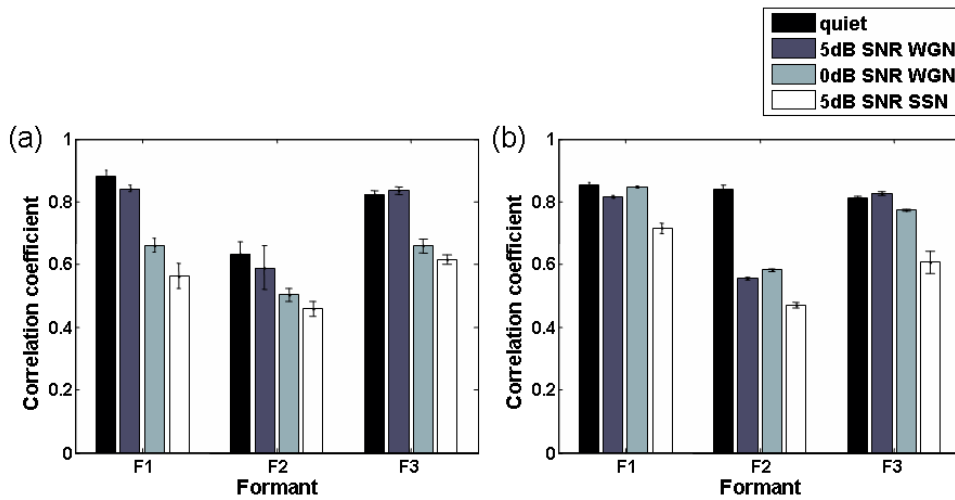


그림 3.7 여러 가지 종류 잡음 하의 평균 포먼트 디코딩 성능변화 (자극율이 5000 pps일 경우). (a)는 시불변-선형 필터뱅크 기반 어음처리방식, (b)는 시변-비선형 필터뱅크 기반 어음처리방식에 대한 평균 포먼트 디코딩 성능

### 3.2 적응효과 적용 전후의 음성시작점 정보 전달특성평가

#### 3.2.1 음성시작점 정보 전달특성평가에 이용된 상세 파라미터

어음처리방식은 시불변-선형 필터뱅크 기반 어음처리방식과 시변-비선형 필터뱅크 기반 어음처리방식을 이용하였으며 각 어음처리방식에 적응효과 적용 전후에 음성시작점 정보 전달특성평가를 수행하였다. 8채널로부터 지정된 어음처리방식에 기반한 전기자극펄스를 생성하였고 자극율은 1250 pps로 설정하였다 [3]. 전기 자극펄스의 최소값은 441  $\mu$ A로 두고 2.5% modulation depth를 갖도록 진폭변조되었다. 음성이 시작된 부분으로부터 20 ms 동안을 시작점으로 가정하였다 [3]. 분류기의 입력으로 채널당 50개의 spike train을 얻어 5 ms time bin을 취하여 계산된 평균 발화율을 이용하였다. 실험에 사용된 잡음 환경은 잡음이 없는 경우와

10, 5 혹은 0 dB SNR 백색잡음 (white Gaussian noise, WGN) 하 그리고 10, 5 혹은 0 dB SNR 음성형태 잡음 (음성과 유사한 형태의 스펙트럼을 갖는 잡음 - speech shaped noise, SSN) 하 등으로 각각의 조건에서 어음처리방식에 적응효과 적용 전후의 성능평가 비교를 수행하였다.

### 3.2.2 적응효과 적용 전후의 음성시작점 정보 전달특성 비교

그림 3.8과 3.9는 어음처리방식에 적응효과 적용 전후의 음성시작점 정보 전달특성을 비교한 대표적인 그림으로 각각 음성 /pa/와 /za/에 대한 결과이다. 잡음이 없는 경우는 두 가지 어음처리방식 모두 적응효과를 적용하지 않아도 음성시작점에 관한 정보의 전달특성은 우수했고 적응효과를 적용할 경우 정보 전달특성은 유사한 수준이거나 우수함을 보였다. 그러나 백색잡음 하에서는 그림 3.8(a),(c)와 그림 3.9(a),(c)에서 관찰되듯이 두 어음처리방식에 적응효과를 적용함에 따라 음성시작점의 정보 전달특성이 향상되었다. 음성형태잡음 하에서도 적응효과 적용 후 정보 전달특성의 향상을 보였으나 그림 3.8(b),(d)와 그림 3.9(b),(d)에서 보이는 바와 같이 일부 잡음레벨에 대하여 감소를 보이기도 하였다. 여러 가지 음성에 대한 결과를 종합하여 보면 잡음이 없을 경우에는 일반적으로 사용되는 시블 변-비선형필터뱅크 기반 어음처리방식을 이용할 경우에도 충분히 우수한 음성시작점 정보 전달특성을 보이며 적응효과를 추가할 경우 전달특성이 유사하거나 약간의 증감을 보였다. 그러나 잡음이 있는 경우 음성시작점의 정보 전달특성이 감소함을 보였다. 잡음 하에서 각 어음처리방식에 적응효과를 적용할 경우 음성의 시작점에서 음성분류성능이 개선되었는데 음성형태 잡음의 경우보다는 백색잡음 하에서 유효하였다.

표 3.1와 3.2는 여러 가지 음성에 대하여 어음처리방식에 적응효과를 적용 전후 음성시작점 정보 전달특성으로 각각 백색잡음 하에서, 음성형태 잡음 하에서의 성능을 나타낸 것이다. 잡음 하에서 적응효과를 적용할 경우 음성시작점 정보 전달특성이 향상되는 경우가 많았으며 특히 백색잡음 하에서 적응효과의 이점이 두드러졌다.



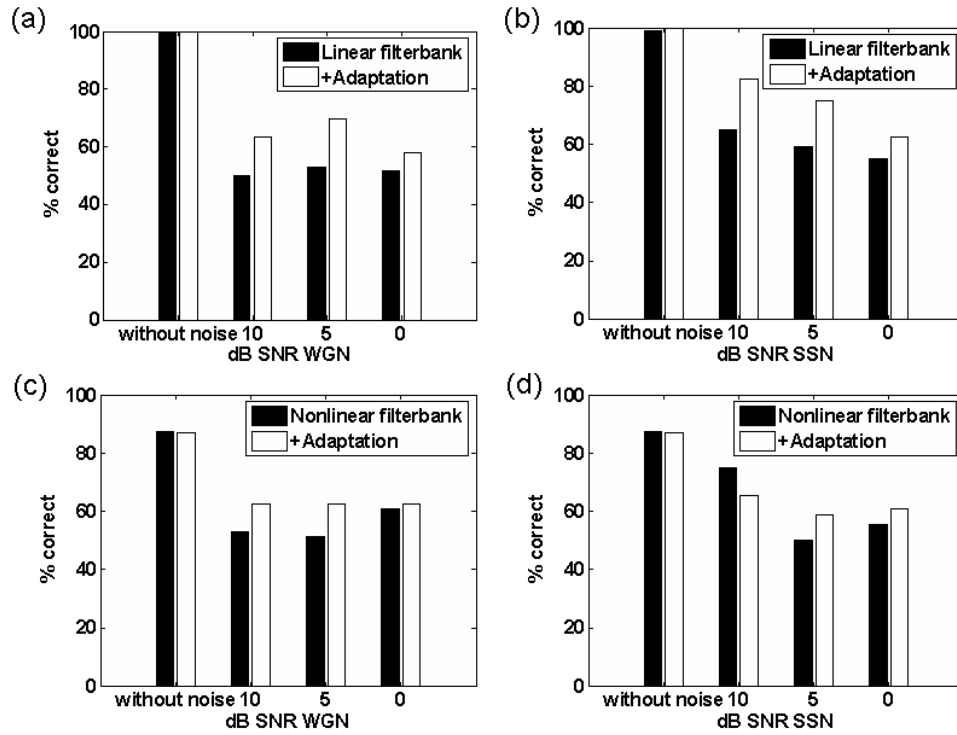


그림 3.8 음성 /pa/를 이용한 경우 어음처리방식에 적응효과 적용 전후의 음성시작점 정보 전달특성 비교. (a),(b)는 시불변-선형 필터뱅크 기반 어음처리방식에 적응효과 적용 전후 비교, (c),(d)는 시변-비선형 필터뱅크 기반 어음처리방식에 적응효과 적용 전후 비교, (a),(c)는 백색잡음 (WGN) 하, (b),(d)는 음성형태잡음 (SSN) 하

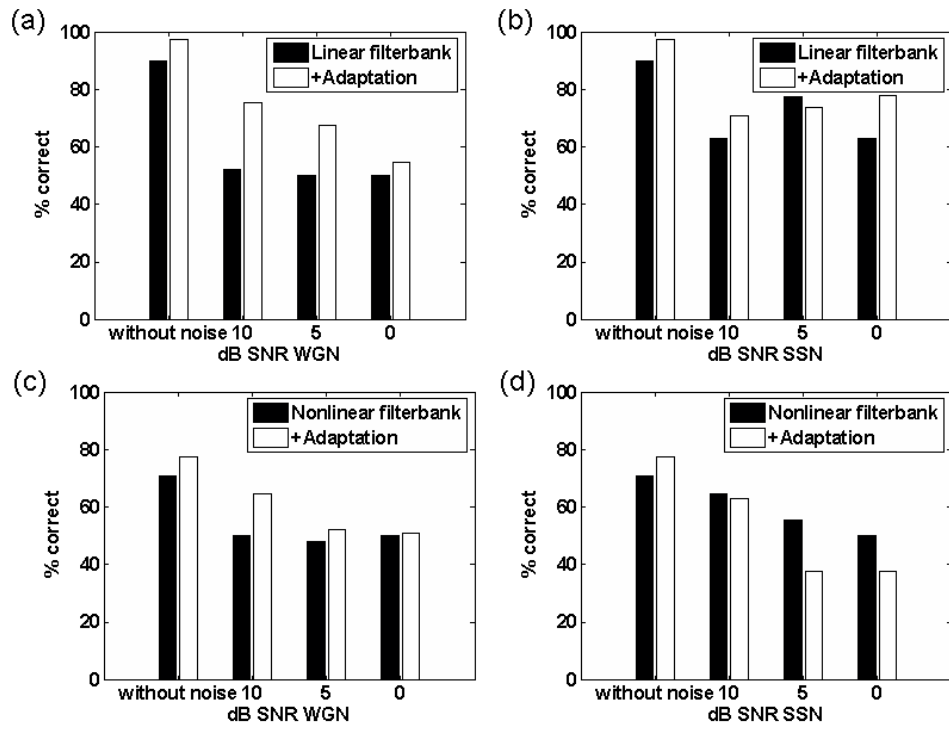


그림 3.9 음성 /za/를 이용한 경우 어음처리방식에 적응효과 적용 전후의 음성시작점 정보 전달특성 비교. (a),(b)는 시불변-선형 필터뱅크 기반 어음처리방식에 적응효과 적용 전후 비교, (c),(d)는 시변-비선형 필터뱅크 기반 어음처리방식에 적응효과 적용 전후 비교, (a),(c)는 백색잡음 (WGN) 하, (b),(d)는 음성형태잡음 (SSN) 하

kind of noise Speech processing strategy	Input speech	without noise		WGN					
				10		5		0	
Linear filterbank/+ adaptation	/ba/	99.38	75	62.5	65	48	65	50	48.75
	/da/	87.5	93.75	62.5	58.75	51.25	50.62	50.62	50
	/pa/	100	100	50	37.5	47.5	50	25	48.13
	/ta/	99.38	96.25	37.5	61.88	62.5	62.5	50	61.25
	/fa/	87.5	71.88	44.37	45	32.5	38.12	38.75	46.25
	/sa/	70	83.75	61.25	81.87	62.5	75	37.5	50
	/va/	100	100	51.25	50	50.62	62.5	62.5	50.62
Nonlinear filterbank/+ adaptation	/za/	89.94	97.48	52.2	75.47	50.31	67.3	50.31	54.72
	/ba/	87.5	85.62	65.63	75.62	62.75	68.1	56.87	61.88
	/da/	71.25	75.62	61.25	48.75	48.75	55	37.5	50
	/pa/	84.38	98.12	51.25	48.13	62.5	75	50	45
	/ta/	80	87.5	58.13	62.5	50	57.5	50	50
	/fa/	73.12	71.88	38	53.75	50	68.13	50	49.38
	/sa/	73.75	87.5	62.5	62.5	54.37	62.5	48.75	62.5
/va/	74.38	87.5	62.5	50	52.5	62.5	37.5	62.5	
/za/	71.07	77.36	50.31	64.78	47.8	52.20	50.31	50.94	

표 3.1 음성시작점 정보 전달특성 (잡음이 없는 경우와 백색잡음 하의 경우). 단위: (%). 회색으로 표기된 부분은 각 어음처리방식에 적응효과를 적용한 경우의 음성시작점에서의 음성유무 분류성능을 나타냄.

Speech processing strategy \ kind of noise	Input speech	SSN					
		10		5		0	
Linear filterbank/+ adaptation	/ba/	87.5	73.12	64.38	73.12	62.5	62.5
	/da/	99.38	78.38	88.12	75	75	63.75
	/pa/	62.5	75	75	62.5	75	74.38
	/ta/	71.25	67.37	73.12	86.25	66.25	75
	/fa/	53.13	51.88	50	55.63	50	53.75
	/sa/	75	62.5	71.88	62.5	56.87	50
	/va/	63.12	62.5	71.88	55.63	54.37	38.12
	/za/	62.89	71.07	77.36	73.58	62.89	78.05
Nonlinear filterbank/+ adaptation	/ba/	66.87	62.5	51.25	58.75	33.12	61.88
	/da/	62.5	56.87	52.5	49.38	66.87	50.62
	/pa/	62.5	62.5	46.88	45	50	57.75
	/ta/	62.5	62.5	63.12	54.37	50	49.38
	/fa/	56.25	60	47.5	48.13	58.13	48.75
	/sa/	81.87	62.5	53.75	61.25	49.38	49.38
	/va/	62.5	62.5	56.25	73.75	50	44.37
	/za/	64.78	62.89	55.35	37.74	50.31	37.74

표 3.2 음성시작점 정보 전달특성 (음성형태 잡음이 하의 경우). 단위: (%). 회색으로 표기된 부분은 각 어음처리방식에 적응효과를 적용한 경우의 음성시작점에서의 음성유무 분류성능을 나타냄.

## 제 4 장 고찰 및 결론

### 4.1 Spike train decoding에 기반한 새로운 어음처리방식 평가방법의 제안

청신경 자극방법을 결정하는 어음처리방식의 성능평가는 인공와우의 성능을 예측하기 위하여 필수적이다 [4-7,21,22]. 많은 연구자료에서 acoustic simulation 등의 방법을 이용하였다. 그러나 실제 이식과 훈련을 거친 후의 어음처리방식의 성능을 효과적으로 예측할 수 있는가에 대한 불확실성이 존재하므로 다각적인 평가를 통한 어음처리방식의 성능검증이 필요하다 [8]. 본 연구에서는 spike train decoding에 기반한 새로운 어음처리방식 평가방법을 제안하였다.

### 4.2 제안된 평가방법을 이용한 모의실험 결과 및 고찰

현재 널리 이용되는 어음처리방식은 입력음성을 주파수 분해하는 주파수 분석부와 주파수 분석결과를 토대로 전기자극펄스의 크기를 결정하는 포락선 (Envelope) 추출부로 구성되어 있다 [15]. 우선 주파수 분석부를 개선한 시변-비선형 필터뱅크 기반 어음처리방식과 널리 이용되는 시불변-선형 필터뱅크 기반 어음처리방식의 성능평가를 수행하였다.

#### 4.2.1. 포먼트 디코딩

여러 가지 잡음 하에서 지정된 어음처리방식에 기반하여 생성된 전기자극펄스에 의하여 자극된 청신경 응답으로부터 spike train decoding 방법을 이용하여 음성 인지에 중요한 역할을 하는 제1, 2 그리고 제3 포먼트 궤적을 추정하여 입력음성 포먼트의 참값과 비교함으로써 어음처리방식의 포먼트 정보 전달특성을 평가하였다 [11,12]. 실험결과 전기자극펄스 자극율의 변화에 상관없이 잡음 하에서 시변-비선형 필터뱅크 기반 어음처리방식의 우수성을 관찰할 수 있었다.

잡음이 없는 경우 혹은 낮은 레벨의 잡음 하의 경우 어음처리방식간의 성능차이를 보기 어렵고 상대적으로 높은 레벨의 잡음 하 혹은 음성 형태의 잡음 하의 경우에도 차이를 관찰할 수는 있으나 얼마나 차이가 나는지 알기 어렵다. 또한 한 회의 학습을 통하여 얻은 디코딩 필터계수와 청신경모델 응답의 통계적 특성에 대하여 독립적인 포먼트 디코딩 성능을 관찰하기 위해서는 여러 회의 시뮬레이션에 대한 포먼트 궤적의 참값과 추정값 간의 유사성을 정량적으로 비교하는 것이 필요하다. 여러 회의 실험을 통한 어음처리방식 성능평가 결과 잡음이 없거나 낮은 SNR의 백색잡음 하의 경우 두 어음처리방식 간 성능은 유사한 수준을 보였다. 이는 일반적으로 사용되는 시불변-선형 필터뱅크기반 어음처리방식을 이용할 경우에도 잡음이 없는 경우에 한하여 인공와우 사용자의 어음 인지율이 높은 것과 연관 지어 생각할 수 있다 [15,21,22]. 그러나 잡음 하에서는 어음 인지율이 감소하며 어음처리방식에 잡음하에서 어음인지에 기여하는 청각계의 비선형성을 더해 줌으로써 어음인지율의 향상이 가능하다 [16,17,21,22]. 실험결과 기존의 시불변-선형 필터뱅크 기반 어음처리방식을 사용했을 경우 높은 크기의 백색잡음 하 혹은 입력음성과 유사한 스펙트럼 특성을 갖는 잡음 하에서는 포먼트 디코딩 성능이 급격히 감소하였고 따라서 낮은 음성인식율이 기대된다. 그러나 시변-비선형 필터뱅크 기반 어음처리방식에 의한 포먼트 디코딩 성능은 0 dB SNR과 같이 높은 크기의 잡음이 존재할 경우에도 완만한 감소를 보이며 시불변-선형 필터뱅크 기반 어음처리방식에 비해 우수한 성능을 보인다. 이는 기저막의 비선형적인 특성을 반영함으로써 잡음에 강인한 특성을 구현할 수 있음을 보여준다.

어음처리방식의 주파수 분석부 즉, 기저막에 해당하는 대역통과필터의 특성은 시간에 따라 고정된 경우가 많으며 이러한 특성은 잡음이 없는 경우에는 큰 문제가 되지 않으나 잡음이 있는 상황에서는 음성의 인식률을 크게 저하시키는 원인이 된다. 정상인의 경우 기저막과 내측유모세포/청신경 시냅스, 외측유모세포 등에서 보이는 비선형성이 음성의 인식에 기여하기 때문에 상당히 큰 잡음 하에서도 성공적인 음성인지에 도움을 줄 수 있다 [16,21,22]. 시변-비선형 필터뱅크 기반 어음처리방식은 이와 같은 비선형성을 추가하였기 때문에 잡음 하에서도 음성 인지에 중요한 포먼트 성분을 청신경 응답으로 충실히 전달할 수 있는 것으로 판단된다.

다.

#### 4.2.2 음성시작점 정보 전달특성 평가

다음으로 어음처리방식에 적응효과를 적용할 경우 음성시작점에 관한 정보의 전달특성을 평가하였다. 정상 청각계에서는 그림 2.6(a)에서 보인바와 같이 귀에 일정 크기의 소리가 인가되었을 때 청신경의 응답은 자극의 시작점에서는 높은 발화율을 보이고 시간이 지남에 따라 감소하는 적응효과가 발생하며 이는 내측유모세포와 청신경 사이의 화학적 시냅스에 기원한다 [3,22-24]. 인공와우 사용자의 경우 소리정보가 전기자극에 의하여 전달되므로 화학적 시냅스에 의한 적응효과를 기대하기 어렵다 [3,22-24]. 본 연구에서는 음성의 시작점을 강조해주는 적응효과를 적용할 경우 시작점 구간에서 음성유무 분류성능을 관찰함으로써 음성의 시작점에 대한 정보의 전달이 더 충실하게 제공할 수 있는지 알아보았다. ‘자음+모음’ 형태의 음성을 이용하여 실험을 수행하였다. 음성이 시작된 이후 20 ms 즉, 음성시작점에서는 주로 자음에 해당하는 부분이 옴으로 여러 가지 자음에 대한 실험을 수행하였다. 모음의 경우 음성파형의 크기가 자음에 비하여 현저하게 크다. 자음 중에서 파열음, 마찰음 등에 대하여 특히 무성음일 경우 음성파형의 크기가 매우 작으므로 이러한 자음으로 시작되는 음성의 경우 시작점의 강조가 필요하다. 그러므로 여러 종류의 자음에 대하여 실험을 수행하였다.

실험결과 잡음이 없는 경우는 적응효과 적용 여부에 상관없이 음성시작점을 잘 검출하였다. 일부 잡음이 없는 경우 적응효과를 적용할 경우 오히려 음성시작점 정보 전달특성이 약간 저하되기도 하였다. 즉, 잡음이 없을 경우에는 적응효과가 추가되지 않은 일반적으로 사용되는 시블변-비선형 필터뱅크 기반 어음처리방식을 이용할 경우에도 효과적으로 음성시작점에 대한 정보가 사용자에게 전달될 것으로 기대할 수 있다. 이는 잡음이 없는 경우에 한하여 인공와우 사용자의 어음 인지율이 높은 것과 연관지어 생각할 수 있다 [15-17,21,22]. 실험에 이용된 세 가지 음성의 경우 여러 가지 잡음 하에서 음성시작점의 정보 전달특성이 저하되었다. 이는 잡음 하에서는 음성시작점에 대한 정보가 충실하게 전달되지 않으며 어

음 인지율이 감소하므로 적응효과를 적용함으로써 음성의 시작점에 대한 정보를 보강해 줄 필요가 있음을 보여준다. 잡음 하에서 각 어음처리방식에 적응효과를 적용할 경우 음성의 시작점에서 음성분류성능이 개선되었는데 음성형태 잡음의 경우보다는 백색잡음 하에서 유효하였다. 이는 어음처리방식에 적응효과를 적용할 경우 음성으로부터 추출된 포락선을 관찰하였을 때 음성형태 잡음 하보다는 백색잡음 하에서 음성의 시작점을 잘 검출하여 강조시켜주기 때문으로 여겨진다. 따라서 음성형태 잡음 하에서도 적응효과의 이점을 충분히 발휘하기 위해서는 적응효과 알고리즘을 개선할 필요가 있다고 판단된다.

#### 4.3 제안된 평가방법의 유용성

잡음이 없는 경우에는 시불변-선형 필터뱅크 기반 어음처리방식을 이용하여도 인공와우 사용자는 충분히 높은 어음 인지율을 보인다 [15-17,21,22]. 또한 어음처리방식에 적응효과를 추가하지 않아도 음성시작점에 대한 정보를 충분히 전달한다. 그러나 잡음하에서는 어음 인지율이 급감한다. 따라서 잡음에 강인한 음성정보 전달특성을 가지는 어음처리방식을 이용하여야 하며 적응효과를 추가함으로써 음성시작점에 대한 정보를 전달하여 줌으로써 음절 간 구분에 도움을 주어야 한다. 이러한 어음처리방식의 개발을 위해서는 실제 인공와우의 이식과 재활훈련에 소요되는 비용과 시간을 감안할 때 모의실험을 통한 평가방법이 필요하며 본 연구에서 제안한 방법이 유용하게 이용될 수 있을 것으로 기대된다.



## 참고문헌

- [1] P. Mitchell, "The prevalence, risk factors and impacts of hearing impairment in an older Australian Community: The Blue Mountains Hearing Study," *XXVI International Congress of Audiology*, Melbourne, Australia, 2002.
- [2] A. Goodsall, N. Condoleon, and R. Cummins, "Replacing Hearing Aids," Analyst Report, Cochlear Ltd., 20 Jun. 2003.
- [3] M. F. Bear, B. W. Connors, and M.A. Paradiso, *Neuroscience: Exploring the brain*, 2nd Ed., Lippincott Williams & Wilkins, 2004, pp.357-385,
- [4] J. T. Rubinstein, "How cochlear implants encode speech," *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 12, no. 5, pp.444-448, 2004.
- [5] P. C. Loizou, "Introduction to cochlear implants," *Tutorial article on cochlear implants that appeared in the IEEE Signal Processing Magazine*, Sept. 1998., pp. 101-130.
- [6] D. B. Grayden, A. N. Burkitt, O. P.Kenny, J. C. Clarey, A. G. Paolini, and G. M. Clark, "A cochlear implant speech processing strategy based on an auditory model," in *Proc. of the 2004, Intelligent Sensors, Sensor Networks and Information Processing Conference*, Dec, 2004., pp.491-296.
- [7] P. J. Blamey, R. C. Dowell, A. M Brown, G. M. Clark, and P. M. Seligman, "Speech processing studies using an acoustic model of a multiple-channel cochlear implant," *J. Acoust. Soc. Am.*, vol. 76, pp.104-110, 1984.
- [8] J. T. Rubinstein, C. Turner, "A novel acoustic simulation of cochlear implant hearing: effects of temporal fine structure," in *Proc. 1st International IEEE EMBS Conference, Neural Engineering*, Mar. 2003, pp.142 - 145.
- [9] K. H. Kim, S. S. Kim, and S. J. Kim, "Improvement of spike train decoder under spike detection and classification errors using support vector machine," *Med. Biol. Eng. Comput.*, vol. 44, no. 1-2, pp.124-30, 2006.

- [10] D. K. Warland, P. Reinagel, and M. Meister, "Decoding visual information from a population of retinal ganglion cells," *J. Neurophysiol.*, vol. 78, pp.2336-50, 1997.
- [11] A. T. Neel, "Formant detail needed for vowel identification," *J. Acoust. Soc. Am.*, vol. 5, pp.125-131, 2004.
- [12] G. E. Peterson, H. L. Barney, "Control methods used in study of the vowels," *J. Acoust. Soc. Am.*, vol. 24, pp.175-184, 1952.
- [13] L. Deng, C. D. Geisler, "A composite auditory model for processing speech sounds," *J. Acoust. Soc. Am.*, vol. 82, pp.2001 - 2012, 1987.
- [14] E. N. Brown, R. E. Kass, and P. P. Mitra, "Multiple neural spike train data analysis: state-of-the-art and future challenges," *Nature neuroscience*, vol. 7, no. 5, pp.456-461, 2004.
- [15] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, and M. Zerbi, "Design and evaluation of a continuous interleaved sampling (CIS) processing strategy for multichannel cochlear implants," *J. Rehabil. Res. Dev.*, vol. 30, no. 1, pp.110-116, 1993.
- [16] C. J. Sumner, L. P. O'Mard, E. A. Lopez-Poveda, and R. Meddis, "A nonlinear filter-bank model of the guinea-pig cochlear nerve: Rate responses," *J. Acoust. Soc. Am.*, vol. 113, pp.3264-3274, 2003.
- [17] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz, "Better speech recognition with cochlear implants," *Nature*, vol. 352, pp.236 - 238, 1991.
- [18] L. M. Litvak, B. Delgutte, and D. K. Eddington, "Improved temporal coding of sinusoids in electric stimulation of the auditory nerve using desynchronizing pulse trains," *J. Acoust. Soc. Am.*, vol. 14, pp.2079-2098, 2003.
- [19] L. M. Litvak, B. Delgutte, and D. K. Eddington, "Auditory nerve fiber responses to electric stimulation: modulated and unmodulated pulse trains," *J. Acoust. Soc. Am.*, vol. 110, pp.368 - 379, 2001.

- [20] I. C. Bruce, L. S. Irlicht, M. W. White, S. J. O'leary, S. Dynes, E. Javel, and G. M. Clark, "A stochastic model of the electrically stimulated auditory nerve: pulse-train response," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 6, pp.630-637, 1999.
- [21] J. H. Kim, D. H. Kim, and K. H. Kim, "A speech processing strategy for auditory prosthesis based on nonlinear filterbank model of biological cochlear," *World congress on Medical Physics and Biomedical Engineering*, Seoul, Korea, 2006.
- [22] 김진호, "말초청각계의 비선형적 모델에 기반한 새로운 인공와우용 어음처리 방식", 연세대학교 대학원, 석사학위논문, 2007년 2월
- [23] L. Geurts and J. Wouters, "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants," *J. Acoust. Soc. Am.*, vol. 105, pp. 2476-2484, 1989.
- [24] R. L. Smith, "Adaptation, saturation, and physiological masking in single auditory-nerve fibers," *J. Acoust. Soc. Am.*, vol. 65, pp. 166-178, 1979.
- [25] S. Kumar, "Neural networks: a classroom approach," McGraw Hill, pp. 137-142, 2005.
- [26] L. Deng, C. D. Geisler, "A composite auditory model for processing speech sounds", *J. Acoust. Soc. Am.* Vol. 82, pp. 2001 - 2012, 1987.
- [27] L. P. Yang, Q. J. Fu, "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," *J. Acoust. Soc. Am.* Vol. 117, pp. 1001 - 1004, 2005.

## *Abstract*

### *Performance Assessment of Cochlear Implant Speech Processing Strategy based on Spike Train Decoding*

Kim, Doo Hee

Dept. of Biomedical Engineering

The Graduate School

Yonsei University

This study suggest a novel method for the evaluation of cochlear implant (CI) speech processing strategy based on neural spike train decoding. From formant trajectories of input speech and auditory nerve responses responding to the electrical pulse trains generated from a specific CI speech processing strategy, optimal linear decoding filter was obtained, and used to estimate formant trajectory of incoming speech. Performance of a specific strategy is evaluated by comparing true and estimated formant trajectories. We compared a newly-developed strategy rooted from a closer mimicking of auditory periphery using nonlinear time-varying filter with a conventional linear-filter-based strategy. It was shown that the formant trajectories could be estimated more exactly in the case of the nonlinear time-varying strategy. The superiority was more prominent when background noise level is high, and the spectral characteristic of the background noise was close to that of speech signals. This confirms the superiority observed from other evaluation methods, such as acoustic simulation and spectral analysis.

Adopting the rapid adaptation, investigated in the responses of auditory nerves, in the cochlear implant speech processing strategy can improve speech intelligibility. This improvement would be mostly due to the better delivery of

place of articulation and speech onset. We applied adaptation in the biological cochlea to speech processing strategies. We could detect the speech onset easier by applying adaptation effect. So, it is possible to expect the improvement of speech intelligibility by offering the information of place of articulation and speech onset to CI users.

---

Keyword: Cochlear implant, Speech processing strategy, Time-invariant linear filterbank, Time-varying nonlinear filterbank, Adaptation, Spike train decoding