

Teager energy operator 기반 순간
주파수 인코딩을 이용하여 개선된
인공와우의 어음처리기

연세대학교 대학원
의 공 학 과
최 성 진

Teager energy operator 기반 순간
주파수 인코딩을 이용하여 개선된
인공와우의 어음처리기

지도 김 경 환 교수

이 논문을 석사 학위논문으로 제출함

2007년 월 일

연세대학교 대학원

의 공 학 과

최 성 진

최성진의 석사 학위논문을 인준함

심사위원 _____ 인

심사위원 _____ 인

심사위원 _____ 인

연세대학교 대학원

2008년 월 일

감사의 글

길다면 길고 짧다면 짧은 2년이라는 시간이 순식간에 지나갔습니다. 지난 2년간 관심 가져주시고, 돌보아주신 모든 분들께 감사의 말씀 드립니다.

부족한 저를 지도해 주시고 이끌어 주신 김경환 교수님께 감사의 말씀 전합니다. 단순히 시간을 보내는 것이 아니라 치열하게 연구하는 법을 배운 소중한 시간이었습니다. 제 평생의 스승이 되실 교수님의 은혜를 잊지 않고 끊임없이 발전해나가는 제자의 모습을 보여드리도록 노력하겠습니다.

부심을 맡아주신 김동윤 교수님, 임창환 교수님께도 감사의 마음을 전합니다. 바쁘신 와중에도 시간을 들여 심사를 해 주셔서 감사드립니다. 학부과정 때부터 많은 가르침을 주신 윤형로 교수님, 이운선 교수님, 이경중 교수님, 윤영로 교수님, 신태민 교수님, 김영호 교수님, 김범민 교수님, 김한성 교수님, 정병조 교수님, 김지현 교수님, 이상우 교수님, 임창환 교수님, 이용흠 교수님 윤대성 교수님, 서종범 교수님, 권기운 교수님께 감사드립니다. 의공학과의 발전을 위해 밤늦은 시간까지 일하시고 연구하시는 모습에서 깊은 감명을 받았습니다.

2년간 동고동락하며 기쁨도, 고난도 함께 나눈 연구실 식구여러분들에게 감사드립니다. 때로는 엄마처럼 때로는 누나처럼 돌보아 주셨던 자현누나, 학부시절부터 많은 것을 가르쳐 주시고 물려주신 저의 사수 진호형, 성실함의 끝을 보여주시고 재미없는 개그에도 항상 웃어주시던 상냥한 상백이형, 중학교 선배이자 고등학교 선배이자 대학교 선배이자 대학원 선배이자 내 친구의 형인 이웃사촌 두희형, 2년간 좋은 것도 나쁜 것도 항상 함께 하며 말 그대로 동고동락한 대학원 동기 진이형, 그리고 길지 않은 시간이었지만 저에게 넓은 세계를 보여주신 종승이형, 새롭게 대학원에 들어와 적응중인 친구 정우에게도 감사의 마음을 전합니다.

학부, 대학원 시절동안 함께 공부한 모든 선배, 동기, 후배 여러분에게 감사드립니다. 희재형, 영진이형, 한정이형, 도원이, 태우, 종기, 태윤이, 종상이, 형일이형, 태영이 등등……. 이 짧은 글로 모든 분들께 감사드리긴 힘들지만 감사하는

마음만은 전달되었으면 합니다.

세상 무엇보다도 소중한 가족들에게도 감사드립니다. 공부한다는 핑계로, 바쁘다는 핑계로 자주 찾아뵙지 못한 점 죄송스럽게 생각합니다. 20살부터 지금까지 6년이라는 긴 시간동안 함께 살지 못하고 떨어져 지내면서 가족의 소중함을 더 크게 느꼈습니다. 항상 저를 생각해주시고 말로 형용할 수 없는 큰 사랑을 베풀어주시는 부모님께 감사드립니다. 그리고 저의 뉘까지 부모님을 모시고 효도한 형에게도 감사의 말씀 드립니다.

마지막으로 직접 말을 전하진 못하지만 도움과 관심을 주신 모든 분들에게 감사의 말씀을 올리며 이만 글을 마칠까 합니다.

차 례

그림 차례	ii
표 차례	iv
국문 요약	v
제1장 서론	1
1.1 연구배경 및 현황	1
1.2 어음처리법의 개발	3
1.3 어음처리기 성능평가방법	5
제2장 방법	8
2.1 어음처리기	8
2.1.1 포락선 이용 어음처리기	8
2.1.2 fine structure 정보를 포함한 어음처리기	10
2.2 포락선 및 순간주파수 추출방법	11
2.2.1 힐버트 변환을 이용한 포락선 및 순간주파수 추출방법	12
2.2.2 TEO를 이용한 포락선 및 순간주파수 추출방법	13
2.3 순간주파수의 양자화	15
2.4 Acoustic simulation을 통한 어음처리기의 성능평가방법	18
2.4.1 멜로디인지실험	18
2.4.2 어음인지실험	19
2.4.3 Acoustic simulation 실험 환경	22
2.4.3.1 어음처리기 비교실험	22
2.4.3.2 연속적인 순간주파수와 양자화된 순간주파수의 비교	24
2.5 힐버트 변환과 TEO의 연산량비교	25
제3장 결과	26
3.1 스펙트로그램 분석 결과	26
3.2 Acoustic simulation을 이용한 청취실험 결과	28
3.2.1 포락선 이용 어음처리기와 순간주파수 적용 어음처리기의 결과	28
3.2.2 양자화된 순간주파수의 영향	32
3.3 힐버트 변환과 TEO의 연산량 비교	34
제4장 고찰 및 결론	35
참고 문헌	38
Abstract	42

그림 차례

그림 1.1 인공와우의 간략도 및 구성요소	2
그림 1.2 인공와우의 block diagram	6
그림 2.1 포락선 정보만을 이용한 어음처리기	8
그림 2.2 fine structure 정보를 포함한 어음처리기	10
그림 2.3 acoustic simulation을 위한 modulation 방법	11
(a) 포락선 이용 어음처리기의 경우	
(b) fine structure 적용 어음처리기의 경우	
그림 2.5 간략화된 DRNL 모델의 간략도 및 파라미터 설정	10
그림 2.4 포락선 및 순간주파수 추출방법	12
(a) 힐버트 변환을 이용	
(b) TEO를 이용	
그림 2.5 포락선과 순간주파수 추출의 예	14
그림 2.6 순간주파수의 양자화	16
(a) 양자화 과정	
(b) 2, 4, 8레벨 양자화의 설명	
그림 2.7 단어 ‘가위’에 대한 순간주파수	17
(a) 연속적인 순간주파수	
(b) 2레벨 양자화	
(c) 4레벨 양자화	
(d) 8레벨 양자화	
그림 2.8 멜로디인지실험	18
(a) 실험에 사용한 6가지 melodic contour	
(b) 용어설명	
(c) Rising-Flat의 스펙트로그램	
그림 3.1 Rising-Flat에 대한 스펙트로그램	26
(a) 입력신호	

(b) 포락선 정보만을 이용한 어음처리기의 결과	
(c) 힐버트 변환 기반 순간주파수 적용 어음처리기의 결과	
(d) TEO 기반 순간주파수 적용 어음처리기의 결과	
그림 3.2 단어 /zai/에 대한 스펙트로그램	· · · · · 27
(a) 입력신호	
(b) 포락선 정보만을 이용한 어음처리기의 결과	
(c) 힐버트 변환 기반 순간주파수 적용 어음처리기의 결과	
(d) TEO 기반 순간주파수 적용 어음처리기의 결과	
그림 3.3 MCI실험의 결과	· · · · · 28
(a) 4채널 어음처리기의 멜로디인지율	
(b) 8채널 어음처리기의 멜로디인지율	
(c) 12채널 어음처리기의 멜로디인지율	
그림 3.4 단어실험의 결과	· · · · · 29
(a) 4채널 어음처리기의 어음인지율	
(b) 8채널 어음처리기의 어음인지율	
(c) 12채널 어음처리기의 어음인지율	
그림 3.5 문장실험의 결과	· · · · · 30
(a) 4채널 어음처리기의 어음인지율	
(b) 8채널 어음처리기의 어음인지율	
(c) 12채널 어음처리기의 어음인지율	
그림 3.6 양자화된 순간주파수로 인한 멜로디 인지율	· · · · · 31
(a) 난이도 ‘어려움’의 결과	
(b) 난이도 ‘쉬움’의 결과	
그림 3.7 양자화된 순간주파수로 인한 멜로디 인지율	· · · · · 33
그림 3.8 힐버트 변환과 TEO의 연산속도	· · · · · 34

표 차례

표 2.1 채널 별 대역통과필터의 중심주파수	9
표 2.2 채널 별 대역통과필터의 3 dB 대역폭	9
표 2.3 한국어문장 리스트	21
표 3.1 양자화된 순간주파수로 인한 멜로디인지율의 p 값	32
표 3.1 양자화된 순간주파수로 인한 어음인지율의 p 값	33

국 문 요 약

Teager energy operator 기반 순간주파수 인코딩을 이용하여 개선된 인공와우의 어음처리기

본 연구는 순간주파수 인코딩을 적용하여 인공와우의 주파수표현성능을 향상시킨 새로운 어음처리기를 제안한다. 또한 본 연구는 입력음으로부터 순간주파수를 추출하기 위하여 Teager energy operator (TEO)를 사용할 것을 제안한다. 스펙트로그램을 통한 시-주파수 분석을 통해 TEO에 기반을 둔 어음처리기가 훌륭한 주파수표현성능이 있음을 보였다. 이와 같이 개선된 주파수표현성능은 힐버트 변환을 이용하여 순간주파수를 추출한 어음처리기에서도 동일하게 나타난다. 포락선과 함께 순간주파수를 사용하는 어음처리기와 기존의 포락선정보만을 이용한 어음처리기의 멜로디인지율과 어음인지율을 알아보기 위해 acoustic simulation을 실시하였다. 그 결과 TEO와 힐버트 변환에 기반을 둔 인코딩방법간의 인지율 차이는 나타나지 않았으며, 두 방법 모두 포락선정보만을 사용한 어음처리기에 비해 향상된 성능을 보였다. 한편 두 순간주파수 추출방법의 연산량비교실험결과 TEO가 힐버트 변환보다 35% 빨랐다.

또한 양자화한 순간주파수를 사용하여도 본래의 연속적인 순간주파수를 사용하는 어음처리기와 비슷한 성능을 가짐을 검증하였다. 양자화한 순간주파수가 어음처리기의 멜로디인지와 어음인지에 영향을 미치는지 알아보기 위하여 acoustic simulation을 실시하였다. 그 결과 두 개의 레벨 이상으로 양자화한 순간주파수는 연속적으로 변화하는 순간주파수에 비해 그리 낮지 않은 인지율을 가짐을 확인하였다.

핵심되는 말: 인공와우, fine structure, 순간주파수, 순간주파수의 양자화, acoustic simulation

제 1 장 서 론

1.1 연구배경 및 현황

정상 청각계에서 소리의 진동은 외이에 위치한 이도 (ear canal)를 통해 고막 (tympanic membrane)으로 전달되고, 이러한 진동은 중이에 위치한 이소골 (ossicles)을 움직인다. 고막의 면적은 이소골의 마지막에 위치한 등골 (stapes)의 면적보다 10 배 정도 크다. 이러한 크기차이로 인해 외이에서 전달된 에너지는 증폭된다. 등골의 움직임은 내이의 난원창 (oval window)을 기계적으로 진동시키고 이 진동은 림프액으로 차 있는 달팽이관 (cochlea)을 자극한다. 림프액의 움직임은 달팽이관에 위치한 기저막 (basilar membrane)을 움직이게 한다. 입력되는 진동의 주파수에 따라 기저막의 각기 다른 지역에서 움직임이 발생한다. 기저막의 base는 좁고 두꺼운 막의 형태를 갖고 있으며 고주파에 반응하고, apex는 상대적으로 넓고 얇은 막의 형태를 가지며 저주파에 반응한다. 기저막 위치에 따라 특성 주파수 (characteristic frequency) 를 가지는 특성을 tonotopy라고 한다. 기저막에는 tonotopy에 따라 유모세포 (hair cell)가 분포하고 있으며 최적주파수별로 하나의 내측유모세포 (inner hair cell; IHC)와 세 개의 외측유모세포 (outer hair cell; OHC)가 위치한다.

기저막의 기계적 움직임은 유모세포의 말단에 위치한 여러 stereocilia들의 움직임에 의해 전기적 형태로 변환된다. Stereocilia들은 얇은 필라멘트로 서로 연결되어 있다. 음파가 입력되어 기저막이 움직이면 stereocilia는 한쪽으로 구부러지면서 이온채널이 열리게 되며, 유입된 이온들에 의해 유모세포의 수용기 전위 (receptor potential)가 발생하며 이 수용기 전위는 유모세포의 시냅스에 전달된다. 유모세포의 시냅스는 spiral ganglion cell에 연결되어 있으며 유모세포의 수용기 전위는 활동전위 (action potential) 형태로 변환된다. Spiral ganglion cell의 축색 다발은 청신경 (auditory nerve)을 이루며 고차 신경계로 정보를 전달하고, 이러한

인코딩 (encoding) 과정에 의해 소리를 인지하게 된다.

뇌수막염, 약물중독 등에 의해 유모세포 (hair cell)가 완전히 손상된 경우 소리가 중이를 통해서 정상적으로 내이에 전달되더라도 유모세포에 있는 stereocilia의 움직임이 일어나지 않으므로 뇌에서는 소리를 인지하지 못한다. 이러한 감각신경성 난청 (sensorineural hearing loss)의 경우, 고막이나 이소골에 손상이 있는 전음성 난청 (conductive hearing loss)과 달리 수술이나 보청기 (hearing aid)를 통해 청력을 회복할 수 없다. 다행히 유모세포가 모두 손상된 경우에도 청신경의 근간을 이루는 spiral ganglion cell은 남아 있는 경우가 많다. 이러한 경우 spiral ganglion cell을 전기적으로 직접 자극하는 방법을 사용하여 청력을 회복시킬 수 있다. 이를 가능케 하는 장치가 인공와우 (cochlear implant)이다.

그림 1.1은 현재 통용되고 있는 인공와우의 구성 요소 및 신호전달 과정을 설명하기 위한 그림이다. 인공와우의 최우선 요소인 마이크 (microphone)는 외부의 소리를 모으는 역할을 하며 소리를 전기적 신호로 전환하는 역할을 한다. 그 후 어음처리기 (speech processor)는 정상 청각계의 기능을 모사하여 전기펄스를 생성한다. 생성된 전기펄스는 피부를 사이로 두고 자기적으로 접합된 내-외부 코일 (external-internal coil)을 통해 인체 내부로 전송된다. 인공와우 사용자의 달팽

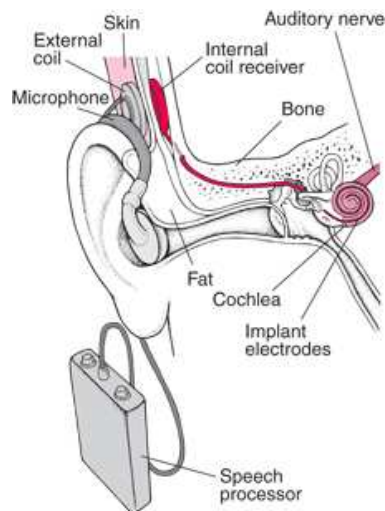


그림 1.1 인공와우의 간략도 및 구성요소

이관 내부에는 청신경을 자극하기 위한 전극이 외과적인 수술로 삽입된다. 전극은 어음처리기에서 생성된 전기펄스로 청신경을 자극한다. 어음처리기는 전기자극펄스를 생성하는 중요한 과정으로 그 구조는 말초청각계 (peripheral auditory system)의 기능에 기인한다. 어음처리기 개선을 통한 인공와우의 성능향상은 추가적인 수술이나, 기계 교체 없이 소프트웨어 개선만으로도 성능을 향상시킬 수 있다는 장점을 가지고 있다 [1]-[3].

1.2 어음처리기법의 개발

현재 사용되는 인공와우 어음처리기의 대부분은 대역통과 필터뱅크 (bandpass filterbank)를 이용하여 입력소리를 몇 개의 주파수 영역으로 분해한다. 이 필터뱅크는 달팽이관내 기저막 (basilar membrane)의 주파수 분해인 tonotopy를 모사한 과정이다 [1]-[3]. 여러 주파수 영역으로 분해된 입력신호는 포락선 추출부 (envelope extractor)를 거치며 이로부터 포락선이 검출된다. 포락선은 음성입력의 여러 가지 성분 중 가장 중요한 어음인지인자로 알려져 있다. 기존 대부분의 어음처리기는 포락선 정보를 이용하여 전기자극펄스를 생성하는데 집중해왔고, 이는 어음 인지에서 훌륭한 결과를 가져왔다. 이 어음처리기들은 사용자에게 따른 차이는 있지만 통상적으로 잡음이 없는 환경에서 높은 어음인지율을 갖게 한다 [1]. 하지만 잡음 하에서는 상대적으로 낮은 인지율을 보이며, 멜로디와 같이 음의 높낮이 인지가 중요한 입력에 대해서는 낮은 성능을 가진다 [4]. 본 연구는 현재의 인공와우 사용자들에게서 나타나는 잡음 하에서 성능감소 및 낮은 멜로디 인지를 해결하기 위한 개선된 어음처리기를 제시하고 평가한다.

기존의 선형필터뱅크에 기반한 어음처리기로는 OHC의 feedback loop에 의한 level-dependent 응답특성이 표현되지 않는다. 지속된 잡음 노출, 이독성 (ototoxic) 약물에 의해 OHC만이 손상을 받은 환자의 경우 다이내믹 레인지가 감소하며, 기저막이 잡음에 민감하게 반응하여서 주파수 분해 능력이 크게 줄어든다. 이러한 사실은 OHC의 중요성을 단적으로 보여준다. Deng 등은 기저막의 비선형

특성을 표현하기 위한 OHC의 damping 파라미터를 달팽이관 모델에 추가함으로써 청신경의 발화 특성이 특히 잡음 하에서 크게 개선됨을 보였다 [5]. 또한 Deng 등은 잡음 하에서 이 비선형 달팽이관 모델을 이용해 얻은 응답으로부터 추출한 모음의 포먼트 (formant)가 선형 달팽이관 모델로부터 얻은 포먼트에 비해 입력음성의 포먼트와 더욱 유사함을 보였다 [6]. 이와 같은 사실은 인공와우의 어음처리기에 OHC의 역할이 적용된 비선형 기저막 모델을 사용함으로써 잡음 하에서의 어음인지 특성을 증가시킬 수 있음을 보여준다. 우리는 Meddis 등이 제시한 비선형 기저막 모델인 dual resonance nonlinear (DRNL) 모델 [7]을 통해 개선된 인공와우의 어음처리기를 제시하였으며, 향상된 어음처리율을 보였다 [8].

신호는 시간에 따라 느리게 변하는 포락선과 빠르게 변하는 fine structure로 나뉜다 [9]. 이 fine structure는 잡음 환경에서 어음인지, 음의 높낮이인지, 소리발생위치 추적, 화자의 성별구분 등의 중요한 역할을 한다고 알려져 있다 [9]. fine structure 정보를 포락선과 함께 인공와우 어음처리기에 적용하기 위한 몇몇 방법들이 제안되어 왔다 [10]-[14]. 이들 중 많은 방법들은 fine structure를 순간위상 (instantaneous phase) 형태로 적용한다 [10]-[12]. 반면에 Zeng 등의 연구자들은 fine structure를 순간주파수 (instantaneous frequency) 형태로 적용할 것을 제안하였고 [13, 14], 잡음 하에서의 음성인지향상을 보임으로서 그 유용성을 증명하였다. Zeng 등은 힐버트변환 (Hilbert transform)을 이용하여 음성에서 포락선과 함께 순간주파수 정보를 추출하였다 [13, 14].

비록 fine structure의 적용을 통한 인공와우의 성능 향상이 실험적으로 증명되었지만 인공와우에서 fine structure를 폭넓게 이용하기 위해서는 몇 가지 고려해야 할 점들이 있다. 순간주파수를 추출하기 위하여 힐버트 변환이 주로 사용되고 있지만, 인공와우의 실시간처리에 적당한 연산량을 가지는지는 아직 검증되지 않았다. Teager energy operator (TEO)는 힐버트 변환처럼 신호에서 fine structure를 분리하기 위해 사용된다 [16, 17]. TEO는 선형 발진기의 에너지를 추적하며 그 연산결과는 신호의 순간진폭 (instantaneous amplitude; envelope)과 순간주파수 (instantaneous frequency; fine structure)의 제곱의 곱 형태이다. Maragos 등은 TEO를 이용하여 순간진폭과 주파수를 추출하는 방법을 제안하였고 [17], TEO를

이용하여 추출한 순간주파수와 포락선을 힐버트 변환으로 추출한 것과 비교하였다 [18].

fine structure를 이용하여 청신경의 생물학적 특성을 적절하게 표현하는 전기 자극펄스를 modulation 하는 방법은 아직 확립되지 않았다. 전기자극펄스를 modulation 할 때 추출한 순간주파수를 바로 사용하기에는 변화하는 속도가 너무 빠르다. 순간주파수의 변화하는 속도를 느리게 하기 위해 여러 방법들이 제안되었다. Zeng등이 제안한 어음처리기는 저역통과필터를 이용하여 순간주파수를 평탄하게 한다. 한편 Collins 등은 소수의 대표주파수를 사용하여 순간주파수를 양자화하였다 [19]. 그들은 양자화된 순간주파수를 사용하여도 본래의 연속 순간주파수에 비해 어음인지율이 크게 떨어지지 않고 본래 포락선만 사용하던 어음처리기에 비해 큰 성능향상이 있음을 보였다.

본 논문의 목적은 TEO 기반 순간주파수 적용 어음처리기를 제안하는 것이다. TEO가 힐버트 변환과 비슷한 성능을 보이면서도 매우 적은 연산량을 가져 실시간처리에 적합하다는 것을 보였다. 힐버트 변환과 TEO의 성능비교를 위하여 멜로디 입력에 대한 각 방법의 인지율을 조사하였다. 멜로디 인지는 fine structure 사용의 가장 큰 장점 중 하나로 알려져 있다. 또한 우리는 fine structure의 또 다른 장점인 잡음 환경에서 어음인지율 향상을 알아보기 위해 단어와 문장입력을 이용한 인지율 실험을 수행하였다. 멜로디와 어음인지율의 측정은 acoustic simulation이라는 정상 청각자를 대상으로 한 청취 실험을 이용해서 이루어진다. 또한 우리는 양자화된 순간주파수가 단순히 어음인지율뿐 아니라 멜로디 인지와 문장을 이용한 어음인지실험에서도 훌륭한 결과를 보이는지 실험하였다. 이를 통해 양자화된 순간주파수의 사용이 유리하다는 것을 증명하였다.

1.3 어음처리기 성능평가방법

소리에 담긴 정보를 얼마나 충실하게 청신경계로 전달하는 전기자극펄스를 생성하는지는 어음처리기에 달려있으며 이에 대한 성능평가는 필수적이다. 최종적으로

로는 인공와우 시술자를 대상으로 어음처리의 성능평가를 수행해야 하지만 이에는 비용, 시간, 대상이 사람이라는 점 등으로 인한 한계점들이 있다. 많은 인공와우 연구에서 어음처리에 따른 청력회복의 정도를 acoustic simulation에 기반을 두어 평가, 측정하고 있다. Acoustic simulation은 특정 어음처리에 기반을 두어 인공와우 시술자에게 들릴 것으로 예상되는 음성을 합성하여 정상청각자에게 들려주어 얻은 어음인지율을 토대로 어음처리를 평가하는 방법이다. 그림 1.2는 전반적인 인공와우 어음처리의 구성도이다 [1]. 그림 1.2에서 실선은 실제 인공와우에 적용되는 어음처리의 역할을 설명한 그림으로 어음처리는 입력신호를 받아서 전기펄스를 생성한다. 이후 생성된 전기펄스는 각 채널의 전극으로 보내져 청신경을 자극한다. 반면에 점선으로 표현된 부분은 acoustic simulation을 이용하여 일반인을 대상으로 한 청취실험의 과정을 보여준다. acoustic simulation에서 각 채널 어음처리의 출력 값은 amplitude modulation되고 모든 채널의 출력 값들이 합산되어 하나의 출력소리가 만들어진다. 각 채널의 출력 값을 amplitude modulation하기 위해 통상적으로 정현파나 좁은 대역의 백색잡음이 많이 사용된다. 본 논문에서 실시된 모든 acoustic simulation은 정현파를 이용하여 modulation 하였다 [2]. 이때 사용하는 정현파는 각 채널 대역통과필터의 중심주파수와 같은 주파수의 신호이다. acoustic simulation에서는 내이에서 전달되는 음성의 시간에 따른 정보가 음성신호의 포락선과 spectral cue에 의해 결정된다는 가정 하에 수행된다. 그러나 음성에는 이 실험에서 고려한 포락선 성분외의 것이

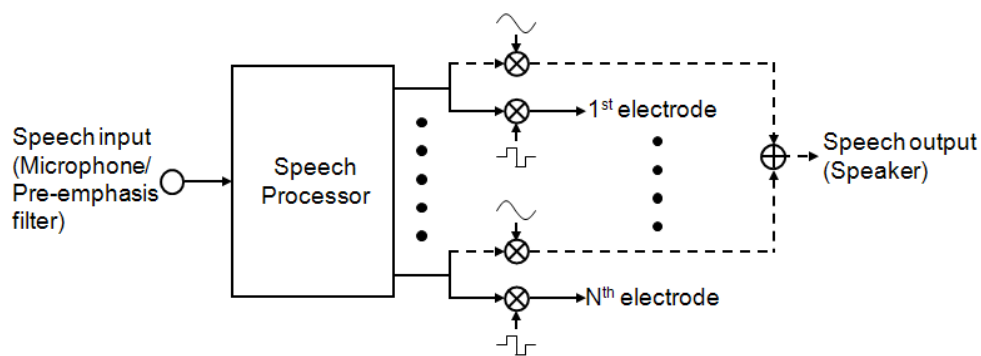


그림 1.2 인공와우의 block diagram

존재한다. 또한 acoustic simulation을 통해서도 영아들의 경우 단 7 채널 인공와우로도 청력회복이 유효하다는 점이나 일부 인공와우 수혜자가 음악을 들을 수 있을 정도의 청력회복을 보이는 경우 등에 대하여는 만족할 만한 설명을 제시하지 못 한다 [20]. 이러한 acoustic simulation의 한계점을 극복하기 위하여 다양한 평가방법을 이용한 검증이 요구된다. 이는 다양한 평가방법을 통하여 우수성이 검증된 어음처리가 실제 인공와우 이식 후에도 우수한 성능을 보일 여지가 많기 때문이다.

Acoustic simulation 이외의 어음처리의 평가방법으로는 청신경의 출력인 spike train을 통해 비교하는 방법이 있다. 비교하고자 하는 어음처리를 통해 전기자극펄스를 생성하고, 그 펄스를 청신경의 입력으로 이용하여 spike train을 만든다. 전기자극펄스를 실제 사람의 청신경에 인가하여 spike train을 얻고 이를 이용하는 것이 이상적이나 어음처리를 평가하기 위해 매번 실험을 재개하는 것은 제약점이 있으므로 Bruce 등이 제시한 청신경 모델을 이용한다 [21]. Bruce는 고양이의 청신경 응답특성을 모사하여 확률적인 청신경 모델을 제시하였다. 실제 청신경이 아닌 모델을 사용하는 또 다른 장점은 청신경의 역치나 불응기 함수 등의 파라미터를 조절할 수 있다는 점이다. 우리는 청신경 모델로부터 얻은 spike train을 이용하여 추정된 포먼트를 원래의 포먼트와 정량적으로 비교함으로써 사용된 어음처리의 성능을 평가하였다 [22]. 이는 입력음성의 특성이 지정된 어음처리에 기반을 둔 전기자극펄스에 의해 청신경의 응답으로 잘 전달될수록 우수한 어음처리라 할 수 있으며 높은 수준의 청력회복을 기대할 수 있다는 가정에 기반을 두고 있다.

제 2 장 방 법

2.1 어음처리기

2.1.1 포락선 이용 어음처리기

그림 2.1은 현재 통용되고 있는 인공와우에서 주로 쓰이는 어음처리기로 음성의 포락선 정보를 바탕으로 출력을 만든다. 이 어음처리기의 첫 번째 단계는 대역통과필터뱅크로 입력 소리를 여러 개의 주파수 대역으로 나눈다. 어음처리기에서 사용하는 대역통과필터뱅크는 고 채널로 갈수록 높은 중심주파수와 대역폭을 가지며 이는 지수적으로 증가한다. 표 2.1과 표 2.2는 실험에서 사용하는 대역통과필터뱅크의 중심주파수와 대역폭이다 [3]. 각 채널의 대역통과필터 출력값은 정류기(rectifier)와 저역통과필터로 이루어진 포락선 추출부에 의해 처리되어 포락선이 된다. 포락선은 시간에 따라 변화하는 입력의 시간정보로 음성인지에서 음성을 구분하는데 가장 중요한 성분이다. 추출된 포락선은 이후 하나의 정현파에 의해 amplitude modulation된다. 이때 사용되는 정현파는 각 채널의 최 앞단인 대역통과필터의 중심주파수와 동일하다 [2].

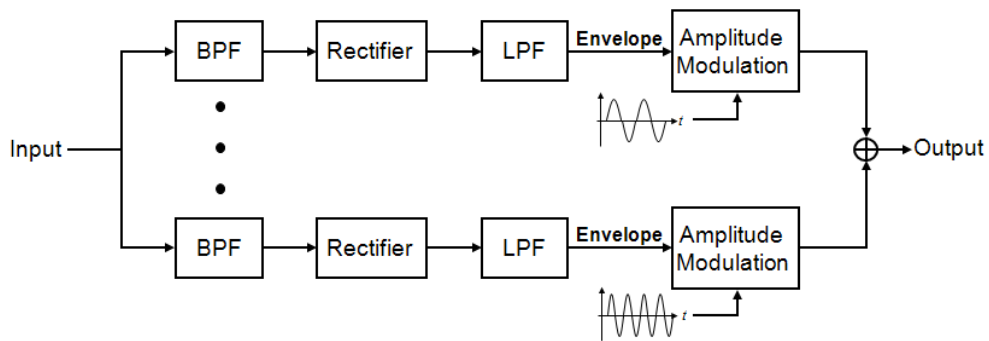


그림 2.1 포락선 정보만을 이용한 어음처리기

표 2.1 채널 별 대역통과필터의 중심주파수 (Hz)

총 채널 수	채널															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	792	3392														
3	545	1438	3793													
4	460	953	1971	4078												
5	418	748	1339	2396	4287											
6	393	639	1037	1685	2736	4444										
8	394	692	1064	1528	2109	2834	3740	4871								
10	322	546	814	1137	1524	1988	2545	3213	4014	4976						
12	274	453	662	905	1190	1521	1908	2359	2885	3499	4215	5050				
16	216	343	486	647	828	1031	1260	1518	1808	2134	2501	2914	3378	3901	4489	5150

표 2.2 채널 별 대역통과필터의 3 dB 대역폭 (Hz)

총 채널 수	채널															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	984	4215														
3	491	1295	3414													
4	321	664	1373	2842												
5	237	423	758	1356	2426											
6	187	304	493	801	1301	2113										
8	265	331	431	516	645	805	1006	1257								
10	204	244	293	352	422	506	607	729	874	1049						
12	165	193	225	262	306	357	416	486	567	661	771	900				
16	120	135	151	170	192	216	242	273	307	345	389	437	492	553	622	700

다음의 수식 (2.1)은 각 포락선이 modulation되는 과정을 설명하는 수식이다.

$$s(t) = \sum_{k=1}^N A_k(t) \cos(2\pi f_k t) \quad (2.1)$$

식 2.1에서 k 는 채널 숫자를, N 은 총 채널수를 가리키고, $s(t)$ 는 어음처리를 통해 합성되는 음성을, $A_k(t)$ 는 k 채널에서의 포락선을 뜻한다. 또한 f_k 는 각 채널 대역통과필터의 중심주파수를 의미한다. 즉 포락선 $A_k(t)$ 는 각 채널의 중심주파수에 해당하는 정현파를 통해 amplitude modulation 된 후 모든 채널의 신호와 더해져 최종적으로 출력된다.

2.1.2 Fine structure 정보를 포함한 어음처리기

그림 2.2는 기존의 어음처리기의 한계점인 잡음환경에서의 인지율저하와 낮은 멜로디인지를 극복하기 위해 새롭게 제안된 어음처리기이다. 본 어음처리기는 이전 모델들이 음성의 포락선정보만을 이용하던 것과 달리 fine structure 정보를 함께 이용한다. 본 모델은 입력음성을 여러 개의 주파수 영역으로 나누는 대역통과필터뱅크를 가지며 이는 그림 2.1의 모델과 동일하다. 그러나 기존의 어음처리기가 대역통과필터의 출력에서 포락선정보만을 추출하던 것과 달리 개선된 어음처리기는 포락선과 함께 순간주파수를 추출한다. 포락선 및 순간주파수를 추출하는 방법에는 힐버트 변환과 TEO가 쓰이며 이 둘에 대한 설명은 다음 단락에서 자세히 하겠다. 추출된 포락선과 순간주파수는 각 채널의 출력을 modulation하는데 사용된다. 이때의 modulation은 포락선 이용 어음처리기의 amplitude modulation과 달리 amplitude-frequency modulation이다. modulation을 위한 공식은 다음과 같다.

$$s(t) = \sum_{k=1}^N A_k(t) \cos \left(2\pi \int_0^t f_{i,k}(\tau) d\tau \right) \quad (2.2)$$

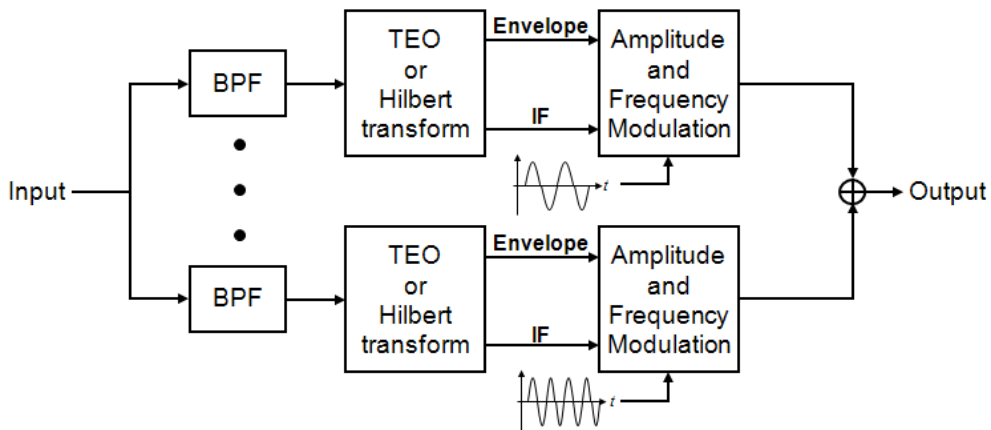


그림 2.2 fine structure 정보를 포함한 어음처리기

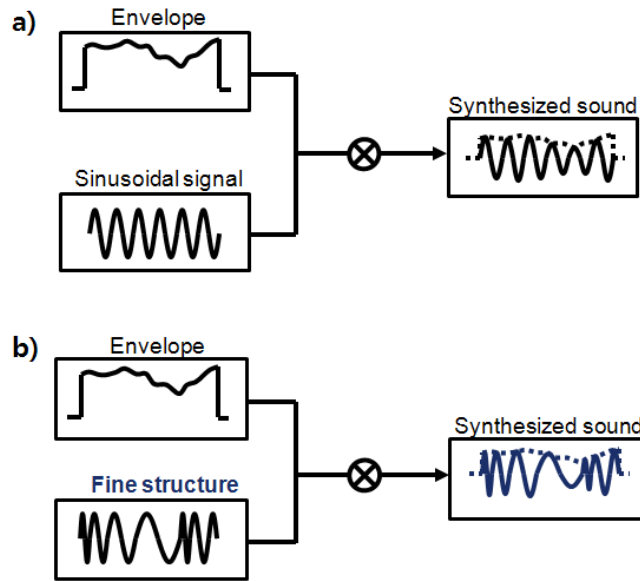


그림 2.3 acoustic simulation을 위한 modulation 방법 a)포락선 이용 어음처리기의 경우
b)fine structure 적용 어음처리기의 경우

식 2.2에서 $A_k(t)$ 는 각 채널의 포락선이고, $f_{i,k}(t)$ 는 순간주파수이다. 이 식에서 순간 주파수를 적분하여 구하는 정현파인 $\cos(2\pi \int_0^t f_{i,k}(\tau) d\tau)$ 를 fine structure라고 한다. 기존의 어음처리기에서의 f_k 와 달리 $f_{i,k}(t)$ 는 시간에 따라 변하는 시간의 함수이다. 이 순간주파수는 각 채널이 가지는 대역을 넘지 않는다.

그림 2.3은 각 어음처리기가 acoustic simulation을 하기 위해 modulation하는 방법을 설명하기 위한 그림이다. a)는 기존의 어음처리기의 경우로 포락선과 하나의 정현파를 이용하여 amplitude modulation 한다. 반면 b)는 새롭게 제안된 fine structure를 이용한 어음처리기로 포락선과 fine structure를 이용하여 amplitude-frequency modulation하여 소리를 생성한다.

2.2 포락선 및 순간주파수 추출방법

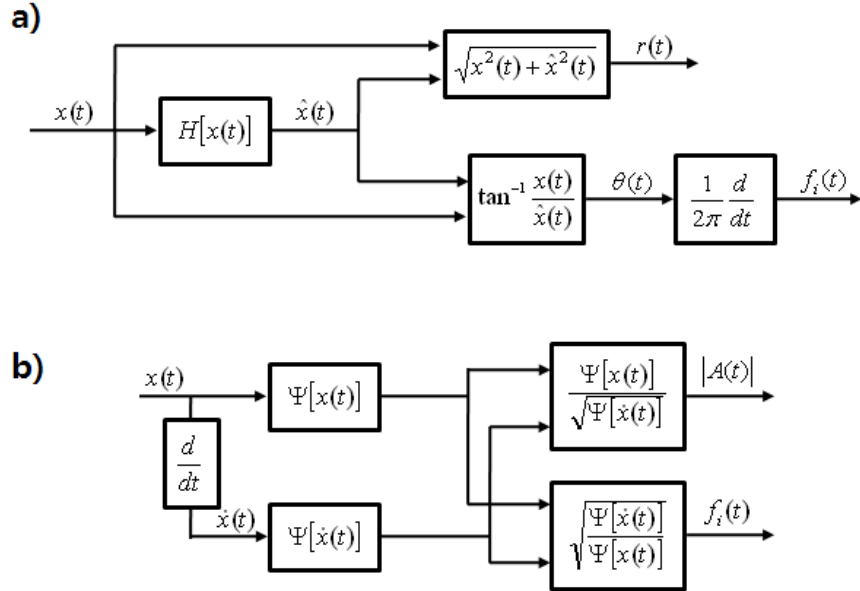


그림 2.4. 포락선 및 순간주파수 추출방법 a)힐버트 변환을 이용 b)TEO를 이용

2.2.1 힐버트 변환을 이용한 포락선 및 순간주파수 추출방법

협대역 (narrowband) 신호 $x(t)$ 는 순간진폭 $A(t)$ 와 순간주파수 $f_i(t)$ 를 이용하여 다음과 같이 표현된다 [15].

$$x(t) = A(t) \cos \left(2\pi f_c t + 2\pi \int_0^t f_i(\tau) d\tau + \theta_c \right) \quad (2.3)$$

식 2.3에서 f_c 와 θ_c 는 채널의 중심주파수와 초기위상이다. 힐버트 변환이 $x(t)$ 에서 $A(t)$ 와 $f_i(t)$ 를 추출하기 위해 이용될 때 $A(t)$ 의 분석적 표현 (analytical representation) 인 $z(t)$ 는 다음과 같다.

$$z(t) = x(t) + j\hat{x}(t) = A(t)e^{j\theta(t)} \quad (2.4)$$

여기서 $\hat{x}(t)$ 는 $x(t)$ 에 대한 힐버트 변환의 결과값이며 $x(t)$ 와 $1/\pi t$ 를 convolution하여 구한다. 순간진폭 $A(t)$ 는 식 2.4에서 바로 구할 수 있으며 순간 주파수인 $f_i(t)$ 는 $\theta(t)$ 를 시간에 대해 미분하여 구한다.

$$r(t) = \sqrt{x^2(t) + \hat{x}^2(t)} \approx |A(t)| \quad (2.5)$$

$$f_i(t) = \frac{1}{2\pi} \dot{\theta}(t) = \frac{1}{2\pi} \frac{d}{dt} \left[\tan^{-1} \left(\frac{\hat{x}(t)}{x(t)} \right) \right] \quad (2.6)$$

식 2.4의 $\hat{x}(t)$ 를 구하기 위해 힐버트 변환을 사용할 때 $x(t)$ 와 $1/\pi t$ 의 convolution 계산은 모든 시간 샘플에서 이루어져야하며 이 과정은 큰 연산량을 유발한다. 그림 2.4의 (a)는 힐버트 변환을 이용하여 포락선과 순간주파수를 추출하는 과정을 설명하는 그림이다.

2.2.2 TEO를 이용한 포락선 및 순간주파수 추출방법

TEO는 힐버트 변환을 대체하여 신호로부터 시간에 따라 변하는 연속적인 순간주파수를 추출하는데 자주 쓰이는 연산 방법이다 [16]. TEO는 신호의 1차 도함수와 2차 도함수를 연산하여 에너지를 구하며, TEO의 연산결과는 신호의 진폭과 주파수 곱의 제곱 형태로 근사된다.

$$\Psi_c[x(t)] = \dot{x}^2(t) - x(t)\ddot{x}(t) \approx [A(t)f_i(t)]^2 \quad (2.7)$$

다음 식은 본래의 신호 $x(t)$ 의 TEO연산 결과인 $\Psi[x(t)]$ 과 $\dot{x}(t)$ 의 결과 $\Psi[\dot{x}(t)]$ 를 이용하여 순간진폭과 순간주파수를 추출하는 방법이다 [17].

$$\frac{\Psi[x(t)]}{\sqrt{\Psi[\dot{x}(t)]}} \approx |A(t)| \quad (2.8)$$

$$\sqrt{\frac{\Psi[\dot{x}(t)]}{\Psi[x(t)]}} \approx f_i(t) \quad (2.9)$$

그림 2.4의 (b)는 TEO를 이용하여 포락선과 순간주파수를 추출하는 과정을 설명

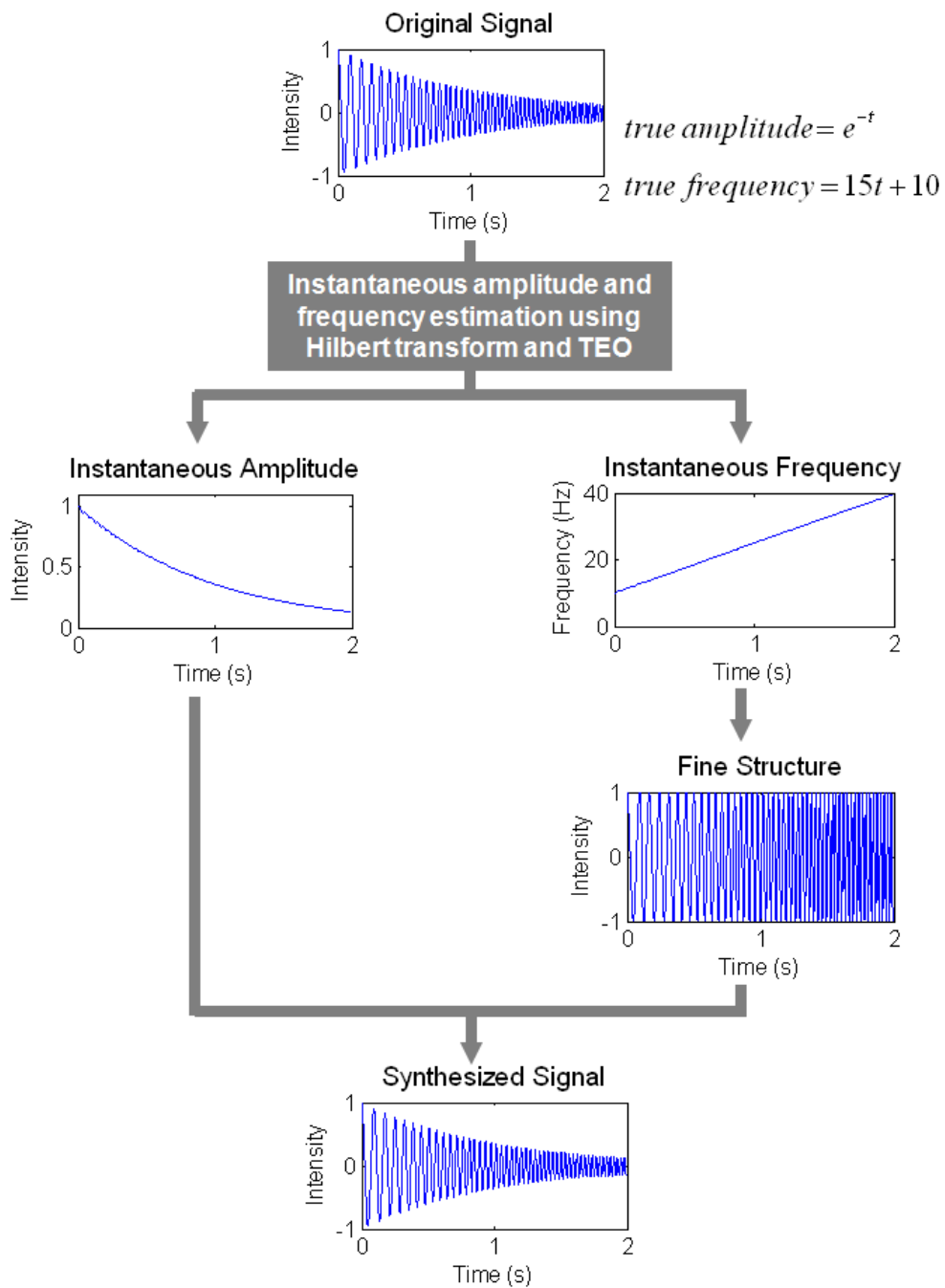


그림 2.5. 포락선과 순간주파수 추출의 예

한 그림이다.

샘플링 된 이산신호에 TEO를 적용할 때 식 2.7의 TEO 연산은 다음과 같이 변형된다.

$$\Psi_d[x(n)] = x^2(n) - x(n-1)x(n+1) \approx [A(n)f_i(n)]^2 \quad (2.10)$$

식 2.10에서 이산신호의 TEO 연산을 위해 3개의 sample 만을 사용하였으며 이때의 연산도 제곱, 곱, 덧셈 등의 간단한 수리 연산일 뿐이다. 이와 같이 TEO는 힐버트 변환에 비해 적은 연산량을 가지며 이는 인공와우 어음처리기의 실시간 연산에 유리하다는 이점을 가진다 [18].

그림 2.5는 힐버트 변환과 TEO를 사용하여 순간진폭 및 순간주파수를 추출하는 예이다. 이때 사용된 입력신호의 진폭은 시간에 따라 지수적으로 감소하는 e^{-t} 이고, 주파수는 선형적으로 증가하는 $15t + 10$ 이다. 추출된 순간진폭은 지수적으로 감소하는 입력신호의 포락선을 충실히 따르며 순간주파수도 2초 시간간격 동안 10에서 40으로 증가하게끔 적절하게 추출되었다. 추출된 두 신호를 이용하여 식 2.2에 따라 amplitude-frequency modulation하면 원본신호가 훌륭하게 복원된다.

2.3 순간주파수의 양자화

힐버트 변환과 TEO를 이용하여 추출한 순간주파수는 시간에 대하여 연속인 신호이며 그 변화속도도 매우 빠르다. 이러한 순간주파수를 인공와우의 어음처리에 적용하기 위해서는 그 변화속도를 줄여야 한다 [19]. 그림 2.6의 (a)는 순간주파수를 양자화하는 과정을 설명하고 있다. 추출된 순간주파수는 저역통과필터를 통하여 느리게 변화한다. 이후 순간주파수는 새롭게 제시되는 양자화과정을 통해 몇 개의 대표 주파수로 양자화된다. 본 실험에서는 연속적인 순간주파수를 2 레벨, 4레벨, 8레벨로 양자화하여 실험한다. 그림 2.6의 (b)는 순간주파수를 양자화할 때 사용되는 주파수를 결정하는 방법을 설명하고 있다. 첫 번째 그림은 한 채널의 전

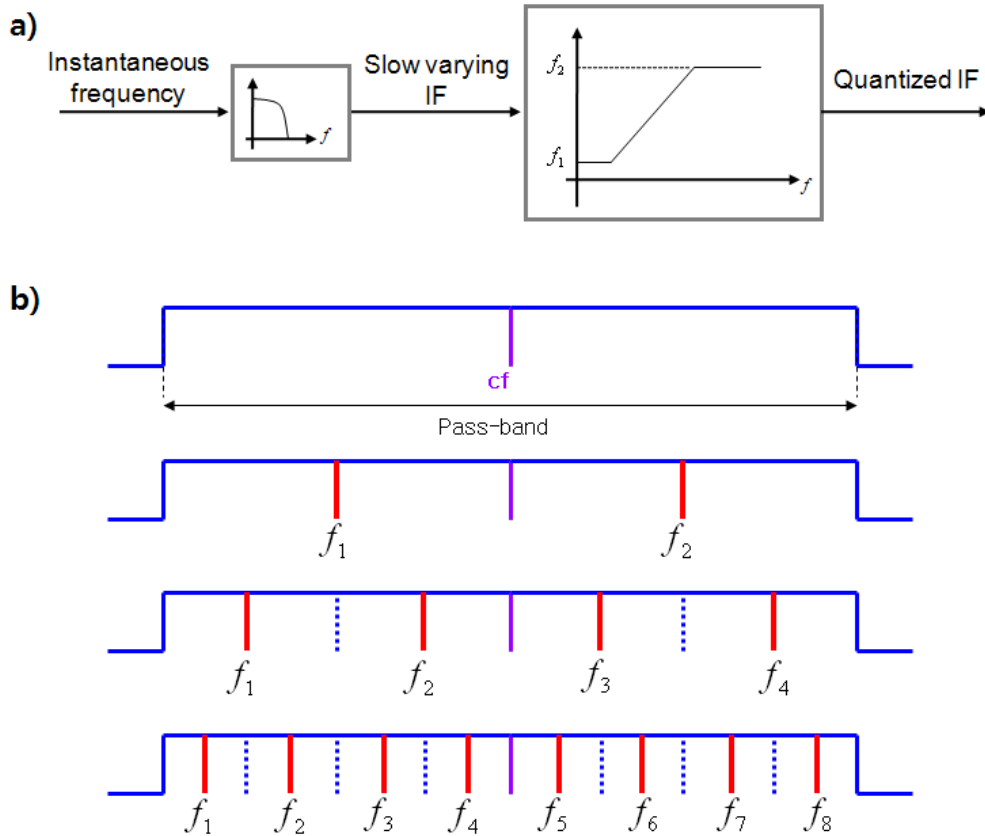


그림 2.6 순간주파수의 양자화 a)양자화 과정 b)2, 4, 8레벨 양자화의 설명.

체 주파수 영역을 보여준다. 두 번째 그림은 전체 주파수 영역을 2개의 레벨로 양자화하는 방법이다. 이때 중심주파수 이하의 모든 주파수는 첫 번째 주파수로 맵핑되고, 중심주파수 이상의 주파수들도 또 다른 주파수로 맵핑된다. 세 번째 그림에서는 4개의 레벨로 양자화하는데 각 채널의 전체 주파수영역을 4개의 subband로 나누며 각 영역에 속하는 주파수는 그 영역의 중간 값으로 맵핑한다. 마지막 그림은 연속적인 순간주파수를 8개의 레벨로 양자화하는 설명이다. 그림 2.7은 4채널 어음처리에 의해 추출한 순간주파수와 그 주파수를 2개, 4개, 8개 레벨로

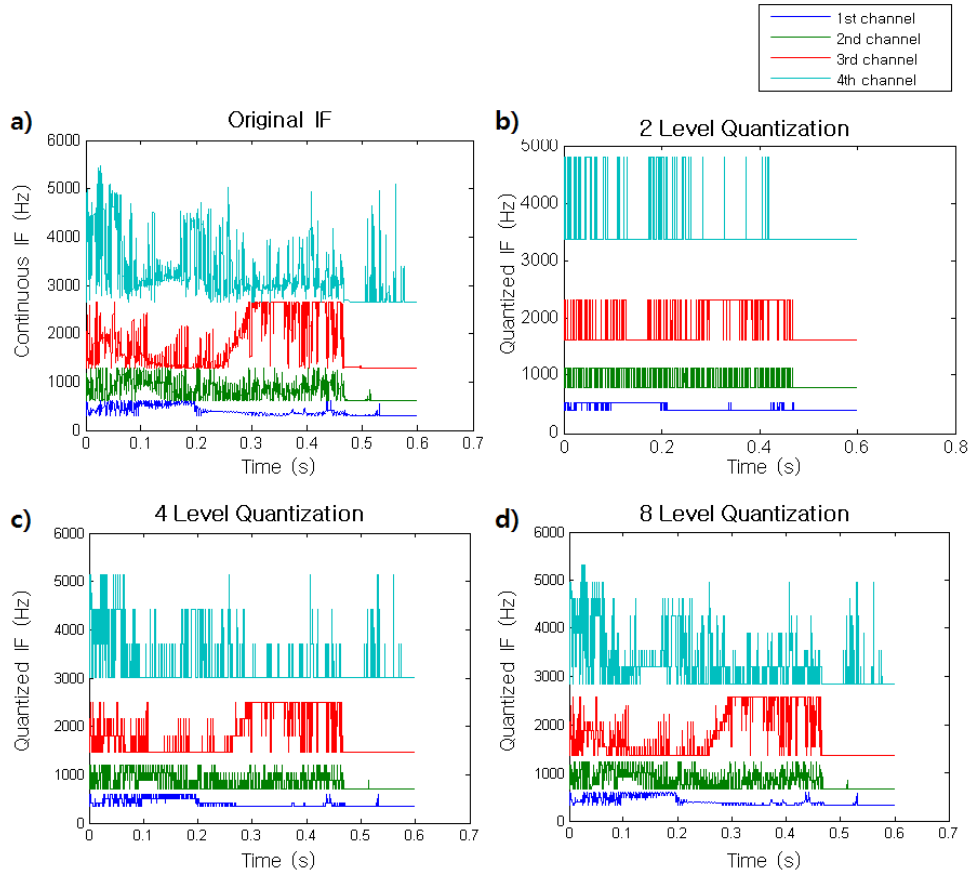


그림 2.7 단어 ‘가위’에 대한 순간주파수 a)연속적인 순간주파수 b)2레벨 양자화 c)4레벨 양자화 d)8레벨 양자화

양자화한 순간주파수이다. 그림 2.7의 (a)는 입력음 ‘가위’의 채널별 연속 순간주파수이다. (b)는 순간주파수를 2개의 레벨로 양자화한 순간주파수로 펄스와 같은 형태를 띄고 있다. (c)는 4레벨로 양자화한 순간주파수의 결과로써 (a)의 연속적인 순간주파수와 유사한 형태를 보인다. 마지막으로 (d)는 8개의 레벨로 양자화한 순간주파수이다. 8개의 레벨로 양자화된 순간주파수는 2레벨이나 4레벨에 비해서 원본 순간주파수와 더욱 비슷하다.

2.4 Acoustic simulation을 통한 어음처리의 성능평가방법

2.4.1 멜로디인지실험

어음처리의 멜로디인지성능을 평가하기 위해 melodic contour identification (MCI) 방법을 채택하였다 [23]. MCI는 본래 인공와우 사용자의 멜로디인지능력을 평가하기 위해 제안된 방법이며 음악 인지와 모음 인지를 예견할 수 있는 실험이

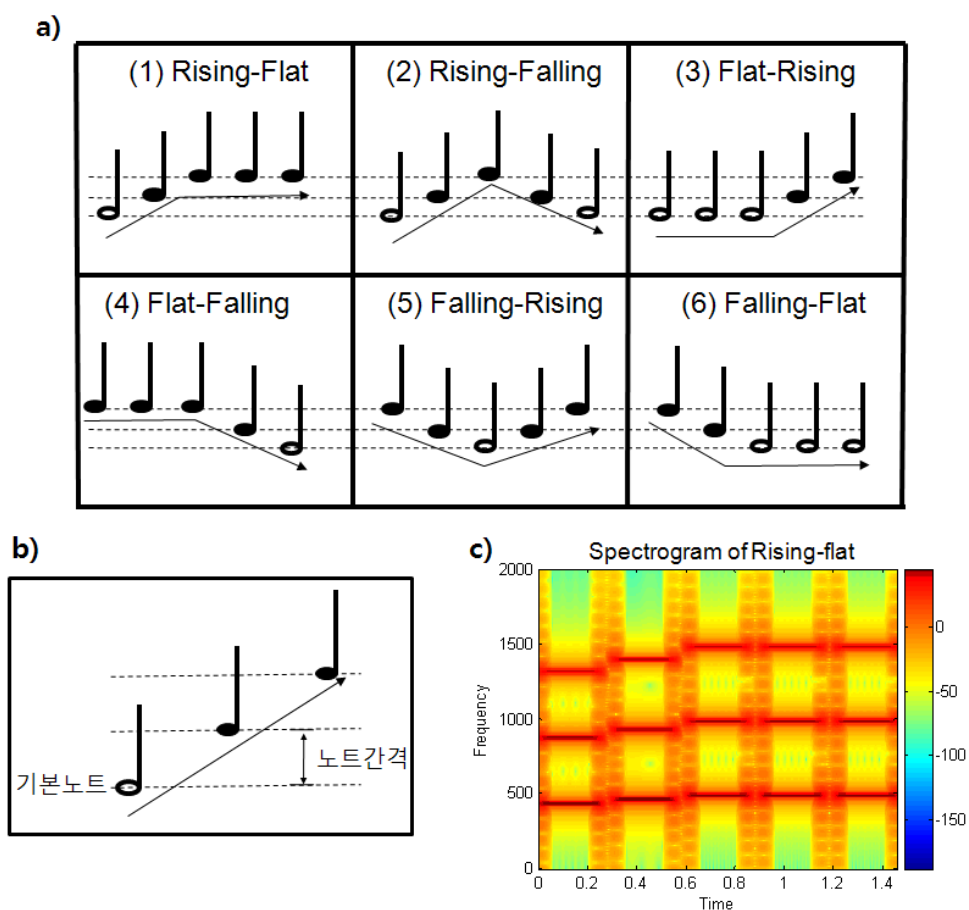


그림 2.8 멜로디인지실험 a)실험에 사용한 6가지 melodic contour b)용어설명
c)Rising-flat의 스펙트로그램

기도하다.

그림 2.8의 (a)는 MCI 실험에서 사용하는 6가지 melodic contour들이다. 그림 2.8의 (b)에서 속이 빈 원으로 표시되는 기본노트는 각 melodic contour가 가진 최저음이다. 또 melodic contour의 각 노트는 일정한 간격을 가지며 이 간격을 조절하여 실험의 난이도를 조절할 수 있다. 각 노트가 가지는 주파수는 공식 $f = 2^{(x/12)} f_{ref}$ 에 따라 결정된다. 이때 x 는 노트와 노트사이에 들어갈 수 있는 가상의 노트개수이며, f_{ref} 는 기본노트이다. 실험에서 기본노트는 220 Hz, 440 Hz, 880 Hz 세 가지로 설정하여 다양한 주파수범위에 대해 표현이 가능하도록 하였다.

그림 2.8의 (c)는 첫 번째 melodic contour인 "Rising-Flat"의 스펙트로그램이다. 기본노트는 440 Hz이며 연속하는 두 노트사이의 가상노트개수는 5개이다. 각 노트는 250 ms의 길이를 가지며 10 ms의 onset과 offset 구간을 가진다. 그리고 노트사이의 시간간격은 50 ms이다. 생성한 melodic contour를 더욱 자연스럽게 만들기 위해 각 노트는 harmonics를 가진다. 노트의 첫 번째 harmonics ($2 \times F_0$)는 fundamental harmonic (F_0)에 비해 -3 dB의 크기를 가지며 두 번째 harmonics ($3 \times F_0$)는 -6 dB의 크기를 가진다.

2.4.2 어음인지실험

어음처리의 음성에 대한 인지율을 실험하기 위해 acoustic simulation을 실시하였다. 어음인지실험은 단어를 이용한 실험과 문장을 이용한 실험으로 나뉜다.

자음-모음-자음-모음 (CVCV) 단어는 모음인지 실험에 사용되는 입력음의 형태이다. CVCV에서 첫 번째 음절과 세 번째, 네 번째 음절은 각각 /s/, /c/, /t/로 고정된다. 반면에 두 번째 음절이자 첫 번째 모음인 /V/는 /a/, /i/, /u/, /e/, /ɔ/로 변화한다. 즉 CVCV 단어에는 5가지 종류인 '사다', '시다', '수다', '세다', '소다'가 있다. 피험자는 단어를 듣고 위의 5가지 종류 중에서 일치하는 것을 찾

아야 한다. 일반적으로 모음인지실험에서 사용하는 단어는 CVC형태이다. 그러나 한국어 단어 중에서는 ‘다’로 끝나는 단어들이 많기 때문에 CVCV 형태를 사용하였다. 5가지 종류인 ‘사다’, ‘시다’, ‘수다’, ‘세다’, ‘소다’는 모두 현실상황에서 빈번히 쓰이는 한국어 단어들이다. 한편 자음인지실험을 위해 모음-자음-모음 (VCV) 형태의 단어가 쓰인다. 본 실험에서는 첫 번째 음절과 마지막 음절을 /ㅏ/로 고정하고 두 번째 음절인 자음들을 /ㄱ/, /ㄷ/, /ㄷ/, /ㄷ/, /ㄷ/로 변화시켰다. 즉 실험에 사용한 다섯 가지 단어는 ‘아가’, ‘아바’, ‘아다’, ‘아마’, ‘아나’이다. 단어를 이용한 어음인지실험의 경우 조용한 환경에서의 어음인지율과 잡음환경에서 인지율을 따로 측정한다. 이때 사용하는 잡음은 acoustic simulation에서 많이 사용하는 음성형태잡음으로 백색잡음을 1200 Hz의 저역통과필터로 처리한 신호이다.

각 어음처리기의 어음인지율을 알아보기 위해 한국어 문장을 사용한 어음인지실험이 추가적으로 실시되었다. 문장실험에서 사용한 한국어문장은 City University of New York (CUNY) 문장과 Hearing in Noise Test (HINT) 문장을 번역하여 녹음하였다. 현재까지 한국어 음성인식실험을 위하여 사용하던 문장들은 문장의 길이도 길고 난이도도 높은 편이었다. 이는 몇몇의 실험에서는 적당할지 몰라도 본 실험에서는 길이가 짧고 난이도도 쉬운 문장을 사용해야 했다. CUNY 문장과 HINT 문장은 인공와우의 어음인지실험에 많이 사용되는 입력으로 쉬운 난이도와 짧은 길이를 가지고 있다. 표 2.3은 어음인지실험에 사용한 문장들의 종류이다. 전체문장은 길이와 난이도가 비슷한 27가지이며, 짧은 문장과 긴 문장 두 종류가 있다. 비록 27가지 문장들의 난이도와 길이가 비슷하지만 각 문장끼리의 차이점은 존재하며 이는 실험결과에도 영향을 미칠 것이다. 이러한 악영향을 줄이기 위해 문장의 순서를 하나씩 시프트해가며 출력음을 생성한다. 따라서 각 피험자는 서로 다른 순서의 문장들을 접하게 된다.

단어실험에서 문장형태의 잡음이 사용되었던 것과 달리 문장실험에서는 또 다른 하나의 문장을 잡음으로 사용한다. 짧은 문장을 이용한 실험에서의 잡음은 ‘옷을 샀습니다’이고, 긴 문장 실험의 잡음은 ‘소파는 이쪽 벽에 놓아 주세요’이다. 단어실험에서 신호 대 잡음비 (SNR)을 조절하듯이 문장실험에서도 신호문장과 잡음문장사이의 크기를 조절해 신호 대 잡음비를 맞춘다. 또한 잡음문장의 길이는 27

표 2.3 한국어문장 리스트

번호	어음인지율 평가를 위한 문장입력	
	짧은 문장	긴 문장
1	식사 하셨어요?	소년이 창문에서 떨어졌다.
2	음식이 비싸다.	고양이가 담을 넘어갔다.
3	드디어 여름이다.	옷은 어디서 세탁하나요?
4	아이가 예쁘다.	이번 주말에 할머니 보러가자.
5	밖에서 먹었어.	나는 약국에서 진통제를 샀다.
6	병원에 가자.	주말까지 편지를 보내세요.
7	눈이 녹았다.	카드로 계산해도 되나요?
8	연못이 더럽다.	먼저 주방부터 치워주세요.
9	공을 던졌다.	아이스크림이 녹고 있다.
10	감기에 걸렸다.	버스가 기차보다 먼저 왔다.
11	딸기잼이 달다.	오늘저녁에 고기 구워먹을래?
12	병이 가득 찼다.	기차가 너무 느려서 늦었다.
13	커피를 마신다.	스키를 타고 점심을 먹읍시다.
14	잔돈 받으세요.	이번추석 어디서 보낼 거야?
15	주사 맞으러 가자	버스가 갑자기 멈췄다.
16	저녁에 어디 갈까?	문에 머리를 부딪쳤습니다.
17	비가 옵니다.	나는 매일아침 커피를 마신다.
18	하늘이 파랗다.	우체부가 편지를 가지고 왔다.
19	잠잘 시간이다.	축구공이 골대에 맞았다.
20	교복을 샀다.	학교가 일찍 끝났다.
21	감자를 길렀다.	어머니가 냄비를 사셨다.
22	냄비가 뜨겁다.	요리사가 빵을 굽고 있다.
23	거울을 본다.	아이들은 딸기를 좋아한다.
24	휴가를 갔다.	개를 데리고 산책을 나갔다.
25	창문이 깨졌다.	어머니가 창문을 닫았다.
26	식탁 위에 있다.	코트는 의자 위에 있다.
27	축구가 끝났다.	신발에 진흙이 묻었다.

가지 신호문장의 길이보다 약간 길다. 그 이유는 잡음문장의 중간부분에 신호문장을 위치시켜 잡음문장이 먼저시작하고 늦게 끝나게끔 제작하였기 때문이다. 이는 실험의 전체 난이도가 증가하는 효과가 있다.

2.4.3 Acoustic simulation 실험환경

Acoustic simulation을 이용하여 인공와우 어음처리기들의 성능을 평가하기 위한 실험을 하였다. 첫 번째 실험은 기존의 어음처리기와 fine structure를 이용하는 어음처리기를 비교하는 것이다. 본 실험에는 순간주파수를 추출하는 두 가지 방법인 힐버트 변환과 TEO의 비교도 포함된다. 두 번째 실험은 양자화된 순간주파수가 어음처리기의 성능에 미치는 영향을 알아보기 위한 실험이다.

Acoustic simulation을 실시하기 위한 음성은 컴퓨터에서 합성하였으며, 피험자는 펜티엄 PC에 장착된 SoundMAX Integrated Digital Audio soundcard의 출력단자에 연결된 Sennheiser HD25SP1 폐쇄형 헤드폰을 통해 합성음을 청취하게 된다. 또한 볼륨은 실험 전 피검자 별로 불편함을 느끼지 않는 레벨로 매 실험마다 조절되었다. 검사에 임하기 전 합성음에 적응하기 위해 충분한 훈련시간이 주어졌다. 피험자는 아이콘을 클릭함으로써 저장된 음성을 들었으며, 어음처리기 및 음성에 대한 정보는 전혀 주어지지 않았다. 합성음이 들려지는 순서는 모든 경우에 대해 임의의 순서로 배열되었다. 문제의 정답에 대한 피드백은 주어지지 않았다.

2.4.3.1 어음처리기 비교실험

기존의 어음처리기와 fine structure 정보를 추가한 어음처리기를 비교하며, 힐버트 변환을 사용한 어음처리기와 TEO를 사용한 것을 비교한다. 각 어음처리기의 멜로디인지율을 알아보기 위해 MCI 실험을 수행하였고, 어음인지율을 알아보기 위하여 단어실험과 문장실험을 실시하였다.

멜로디실험을 위해 사용된 어음처리기의 채널수는 4, 8, 12 채널이었으며

melodic contour의 종류는 그림 2.8 (a)의 6가지였다. 실험에서는 220 Hz와 880 Hz의 기본노트를 사용하여 다양한 주파수범위의 멜로디변화를 실험하였다. 또 연속하는 두 노트 사이의 가상노트 수를 1개, 3개, 5개로 정하여 실험난이도를 3종류로 나누었다. 1개의 가상노트는 노트들의 변화폭이 적어 난이도가 ‘어려움’인 실험이었고, 각각 3개는 ‘중간’, 5개는 ‘쉬움’의 난이도를 가진다. 비교하기 위한 어음처리기는 포락선 정보만을 이용한 기존의 어음처리기와 힐버트 변환에 기반을 둔 fine structure 적용 어음처리기, TEO 기반 fine structure 적용 어음처리기 3종류이다. 따라서 각 피험자는 총 324개의 출력음 (6 melodic contour X 2 기본노트 X 3 난이도 X 3 채널 수 X 3 어음처리기) 을 듣고 6가지의 melodic contour 중에서 일치하는 것을 고르게 된다. 본 실험에 참여한 피험자는 총 12명 (남성 11명, 여성 1명)이며 젊은층의 정상청각자이다. MCI 실험을 통해 비교하기 위한 3 가지의 어음처리기들을 사용하여 소리들을 합성한다. 합성된 소리는 6가지의 melodic contour 중의 하나이다. 피험자가 마우스로 클릭함으로써 합성한 소리가 재생되고, 피험자는 본인이 들은 소리를 6가지의 melodic contour 중에서 가장 가까운 소리를 선택하게 된다. 멜로디 변화를 잘 표현하는 어음처리기는 6가지의 원본 melodic contour와 비슷한 소리를 생성한다. 반면에 멜로디 변화를 잘 표현하지 못하는 어음처리기는 소리의 변화가 잘 구분되지 않는 소리를 생성하게 되어 피험자가 맞는 소리를 선택하는데 어려움을 겪게 된다.

단어입력을 이용한 어음인지실험에는 15명(남자 9명, 여자 6명)의 피험자가 참여하였다. 성능평가를 위하여 실험에 사용된 어음처리기는 멜로디인지실험과 동일한 3가지이며, 4, 8, 12채널의 채널수를 가진다. 실험의 잡음환경은 SNR이 무한대인 조용한 환경과 2.5 dB와 1.5 dB의 음성형태잡음이 사용된다. 입력단어는 5종류의 CVCV와 VCV이며 각 단어는 한 명의 남성 화자에 의하여 녹음되었다. 따라서 각 피험자가 듣게 되는 출력음은 360개 (10 입력음 X 3 채널수 X 3 잡음조건 X 3 어음처리기)이다. 실험은 180개의 CVCV를 이용한 실험과 180개의 VCV를 이용한 실험으로 나뉘어서 진행되었으며 피험자는 두 실험 사이에 충분한 휴식시간을 제공받았다. 피험자는 헤드폰으로 출력음을 듣고 5개의 보기 중에서 맞는 것을 고르게끔 지시받았다.

문장입력을 이용한 어음인지실험에는 27명(남자 15명, 여자 12명)의 피험자가 참여하였다. 실험에는 4, 6, 8 채널의 어음처리가 사용되었으며 신호문장과 잡음 문장의 크기 비인 SNR은 5 dB, 2 dB, 0 dB 였다. 전체 27가지 조건 (3 채널 수 X 3 잡음조건 X 3 어음처리)이 27개의 입력문장과 일대일 대응하여 27가지의 출력음을 만든다. 앞에서 말한바와 같이 27개 입력문장은 하나씩 시프트 하여 또 다른 27개의 출력음을 만든다. 따라서 각각 27개의 문장으로 이루어진 27개의 문장세트가 만들어져, 각 피험자마다 다른 문장세트로 실험을 하였다. 피험자는 헤드폰으로 문장을 듣고 들리는 문장을 받아쓰기한다. 피험자가 실험을 하는 동안 잡음문장인 ‘옷을 샅습니다’와 ‘소파는 이쪽 벽에 놓아주세요’는 항상 제시되어 언제 든지 확인할 수 있게 하였다. 피험자는 제시된 잡음 문장을 보고 섞여있는 문장을 받아쓰기하였다. 각 피험자는 짧은 문장에 대한 실험과 긴 문장에 대한 실험을 두 번 실시하였다.

2.4.3.2 연속적인 순간주파수와 양자화된 순간주파수의 비교

양자화된 순간주파수가 어음처리의 성능에 영향을 주는지 알아보기 위하여 본래의 순간주파수를 이용한 어음처리기와 비교하였다. 성능확인을 위해 MCI를 이용하여 멜로디인지를 실험하였고, 문장입력을 이용하여 어음인지를 실험하였다.

10명의 피험자를 대상으로 멜로디실험이 진행되었다. 멜로디실험을 위해 사용된 어음처리기의 채널수는 4, 8, 12 채널이었다. 실험에 사용한 기본노트는 220 Hz 이었으며 연속하는 두 노트 사이의 가상노트 수를 1개와 3개로 결정하였다. 1개의 가상노트를 사용한 경우 난이도를 ‘어려움’으로 설정하고 3개를 사용한 경우 ‘쉬움’으로 설정하였다. 비교하기 위한 어음처리기는 포락선 정보만을 이용한 기존의 어음처리기와 TEO 기반 fine structure 적용 어음처리기, 양자화된 순간주파수를 이용한 어음처리기였다. 순간주파수를 각각 2개, 4개, 8개의 레벨로 양자화하였다. 따라서 각 피험자는 총 180개의 출력음 (6 melodic contour X 3 채널 X 1 기본노트 X 2 난이도 X 5 어음처리) 을 듣고 6가지의 melodic contour 중에서 일치하는 것을 고르게 된다.

멜로디실험과 동일한 10명의 피험자들이 문장을 이용한 어음인지실험에 참여하였다. 실험에는 4, 8, 12 채널의 어음처리가 사용되었으며 SNR이 2 dB인 한 가지 잡음조건만 사용되었다. 전체 15가지 조건 (3 채널 수 X 1 잡음조건 X 5 어음처리)이 15개의 입력문장과 일대일 대응하여 15가지의 출력음을 만든다. 이전 문장실험과 같은 방법으로 15개의 문장세트가 생성되었다. 각 피험자마다 15개의 문장세트 중 무작위로 선택된 하나의 문장세트에 의해 실험을 한다.

2.5 힐버트 변환과 TEO의 연산량비교

앞서 말한바와 같이 힐버트 변환이 TEO보다 연산량이 많음을 밝히기 위한 실험을 계획하였다. 실험은 10초의 길이를 가지는 하나의 정현파를 각 방법이 처리하는데 걸리는 연산 시간을 측정하는 것이다. 실험에는 그림 2.2의 어음처리가 사용되었으며 순간주파수를 추출하는 방법은 힐버트 변환과 TEO였다. 시간 측정을 위해 사용한 어음처리의 채널수는 16 채널이었다. MATLAB을 이용하여 정현파 입력을 load한 cpu time과 어음처리를 통해서 신호를 재합성하고 save한 cpu time의 차이를 연산시간으로 하였다. 이와 같은 연산시간 측정 실험을 40회 반복하여 평균 연산시간을 계산하였다.

제 3 장 결 과

3.1 스펙트로그램 분석 결과

그림 3.1의 (a)는 6가지의 melodic contour 중 첫 번째인 “Rising-Flat”의 스펙트로그램이다. (b)는 포락선정보만을 이용하여 amplitude modulation하여 얻은 결과 스펙트로그램으로 8채널 어음처리를 사용하였다. (c)는 힐버트 변환을 이용하여 순간주파수를 추출하고 포락선과 함께 modulation 하여 얻은 결과이며, (d)는 TEO를 이용하여 추출한 결과이다. (b)의 포락선만을 이용한 어음처리에 의

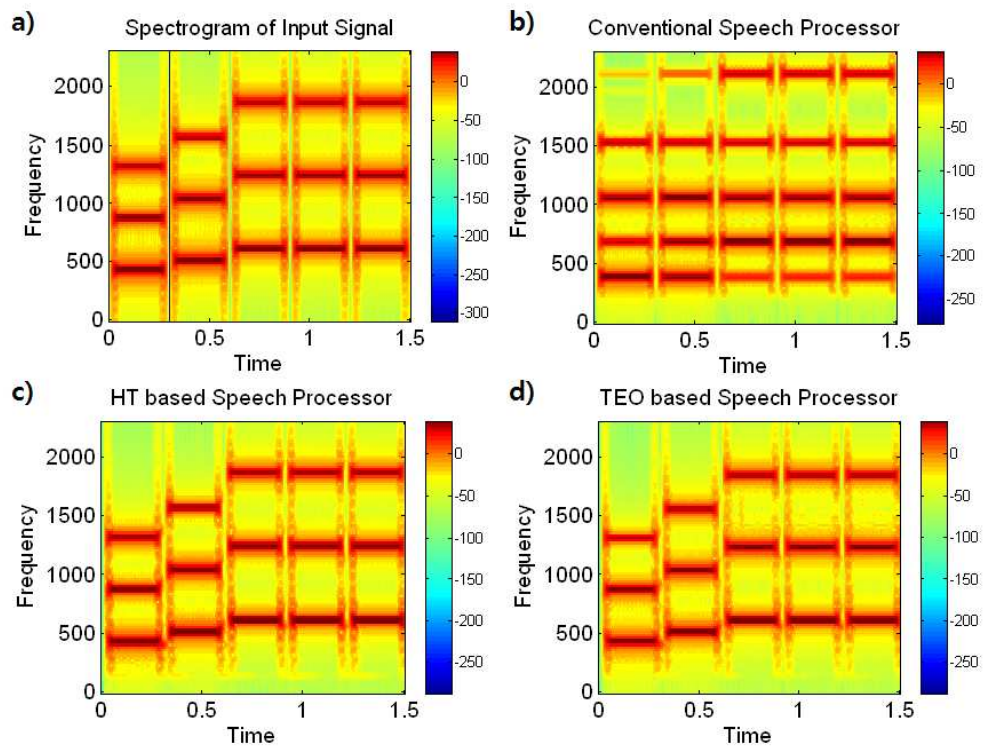


그림 3.1 melodic contour “Rising-Flat”에 대한 스펙트로그램 a)입력신호 b)포락선 정보만을 이용한 어음처리의 결과 c)힐버트 변환 기반 순간주파수 적용 어음처리의 결과 d)TEO 기반 순간주파수 적용 어음처리의 결과

한 결과는 입력의 스펙트로그램과 크게 차이가 나며, 순간주파수를 이용한 결과인 (c)와 (d)는 입력인 (a)와 매우 유사하다.

그림 3.2의 (a)는 단어 /zai/의 스펙트로그램이다. /zai/에서 자음 a가 i로 변하는 순간에서는 입력의 주파수변화가 극심하다. 이와 같은 포먼트 변화를 잘 표현하는 어음처리기는 높은 어음인지율을 가질 것으로 예상된다. (b)는 포락선정보만을 이용하여 amplitude modulation하여 얻은 결과 스펙트로그램으로 8채널 어음처리기를 사용하였다. (c)는 힐버트 변환을 이용하여 순간주파수를 추출하고 포락선과 함께 modulation 하여 얻은 결과이며, (d)는 TEO를 이용하여 추출한 결과이다. 그림 3.1의 멜로디입력을 이용한 스펙트로그램 결과와 마찬가지로 (b)의 포락선

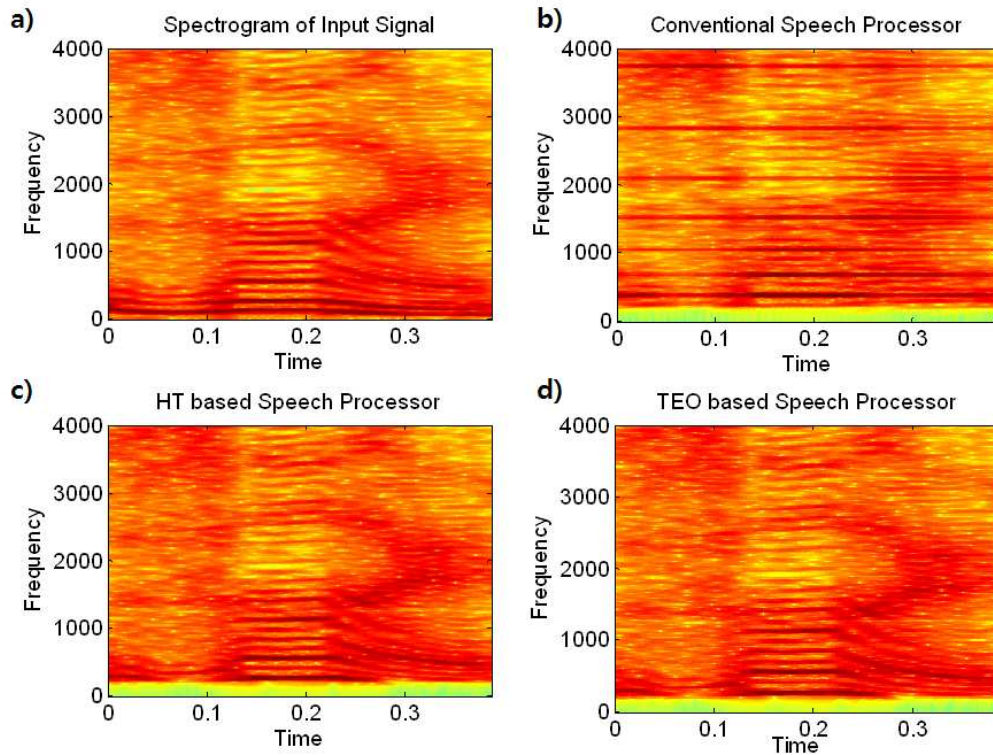


그림 3.2 단어 /zai/에 대한 스펙트로그램 a)입력신호 b)포락선 정보만을 이용한 어음처리기의 결과 c)힐버트 변환 기반 순간주파수 적용 어음처리기의 결과 d)TEO 기반 순간주파수 적용 어음처리기의 결과

정보만을 이용한 어음처리에 의한 결과는 입력의 스펙트로그램과 크게 차이가
 나며, 순간주파수를 이용한 결과인 (c)와 (d)는 (a)와 매우 유사하다.

3.2 Acoustic simulation을 이용한 청취실험 결과

3.2.1 포락선 이용 어음처리기와 순간주파수 적용 어음처리기의 결과

그림 3.3은 포락선정보만을 이용한 인공와우의 어음처리기와 fine structure정
 보를 추가하여 신호를 처리하는 어음처리기간의 멜로디인지율이다. a)는 4채널 어

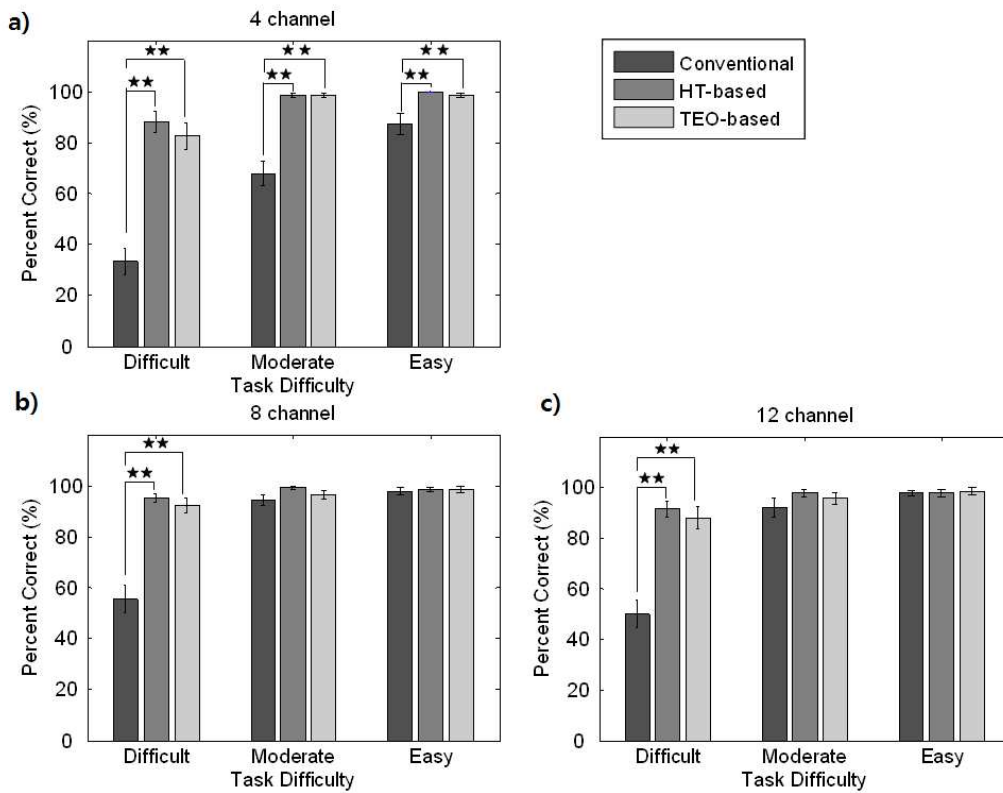


그림 3.3 MCI실험의 결과 a)4채널 어음처리기의 멜로디인지율 b)8채널 어음처리기의
 멜로디인지율 c)12채널 어음처리기의 멜로디인지율

음처리기의 멜로디인지율을 실험의 난이도에 따라 그린 그래프이다. 결과를 보면 모든 난이도에서 fine structure를 적용한 새로운 어음처리기의 멜로디인지율이 포락선 정보만을 이용한 어음처리기의 인지율보다 크게 증가하였다. 순간주파수 적용 어음처리에 의한 멜로디인지율의 증가는 통계적으로 유의미하였다 (one-way ANOVA, $p < 0.01$). 또한 순간주파수를 추출하는 두 방법인 힐버트 변환과 TEO로 인한 멜로디인지율의 차이는 없었다 (t-test, difficult: $p = 0.071$, moderate: $p = 0.999$, easy: $p = 0.166$). 8 채널과 12 채널 어음처리기의 인지율결과 (그림 3.3 (b)와 (c))는 난이도가 ‘어려움’인 경우에만 통계적으로 유의미한 향상이 보였다.

그림 3.4는 단음절 단어를 사용하여 실험한 어음인지율결과이다. (a)는 4채널 어음처리기의 어음인지율을 SNR에 대해 그린 그래프이다. SNR이 2.5 dB와 1.5

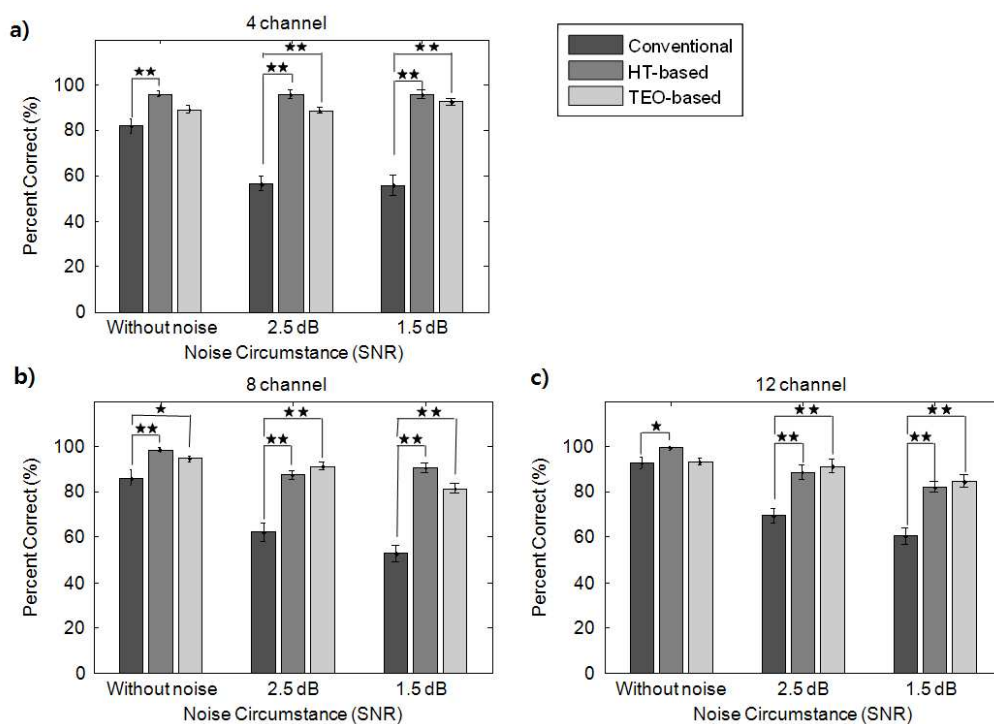


그림 3.4 단어실험의 결과 a)4채널 어음처리기의 어음인지율 b)8채널 어음처리기의 어음인지율 c)12채널 어음처리기의 어음인지율

dB인 잡음이 있을 때 fine structure를 적용한 새로운 어음처리기의 어음인지율은 포락선 이용 어음처리기의 인지율에 비해 유의미한 증가를 보인다 (one-way ANOVA, $p < 0.01$). 이와 같은 결과는 8 채널, 12 채널로 채널이 증가하였을 때도 발견된다. 멜로디인지율결과와 마찬가지로 잡음이 있을 때 힐버트 변환과 TEO로 인한 어음인지율 차이는 없었다.

그림 3.5는 한국어문장을 사용하여 어음처리기들의 어음인지율을 조사한 결과이다. (a)는 4채널 어음처리기를 사용했을 때의 어음인지율을 잡음의 SNR에 대해 그린 그래프이다. 그래프를 보면 포락선 정보만을 이용한 어음처리기에 비해 fine structure를 이용하는 새로운 어음처리기의 어음인지율이 항상 높았다. 또한 잡음이 0 dB로 아주 심한 경우에도 포락선 이용 어음처리기가 낮은 인지율을 보이는 것에 비해 순간주파수 적용 어음처리기는 60% 정도의 어음인지율을 보였다. 이와

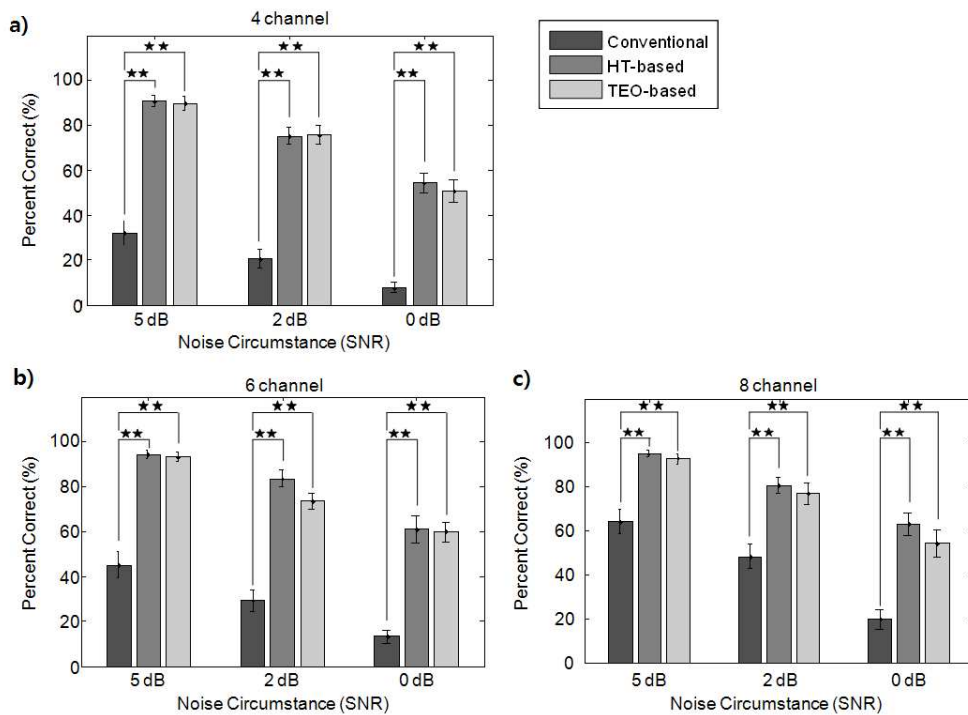


그림 3.5 문장실험의 결과 a)4채널 어음처리기의 어음인지율 b)8채널 어음처리기의 어음인지율 c)12채널 어음처리기의 어음인지율

같은 결과는 6 채널, 8 채널로 채널이 증가하였을 때도 유지된다 (그림 3.5 (a)와 (b)). 모든 채널과 잡음환경에서 힐버트 변환과 TEO에 의한 어음인지율 차이는 발견되지 않았다.

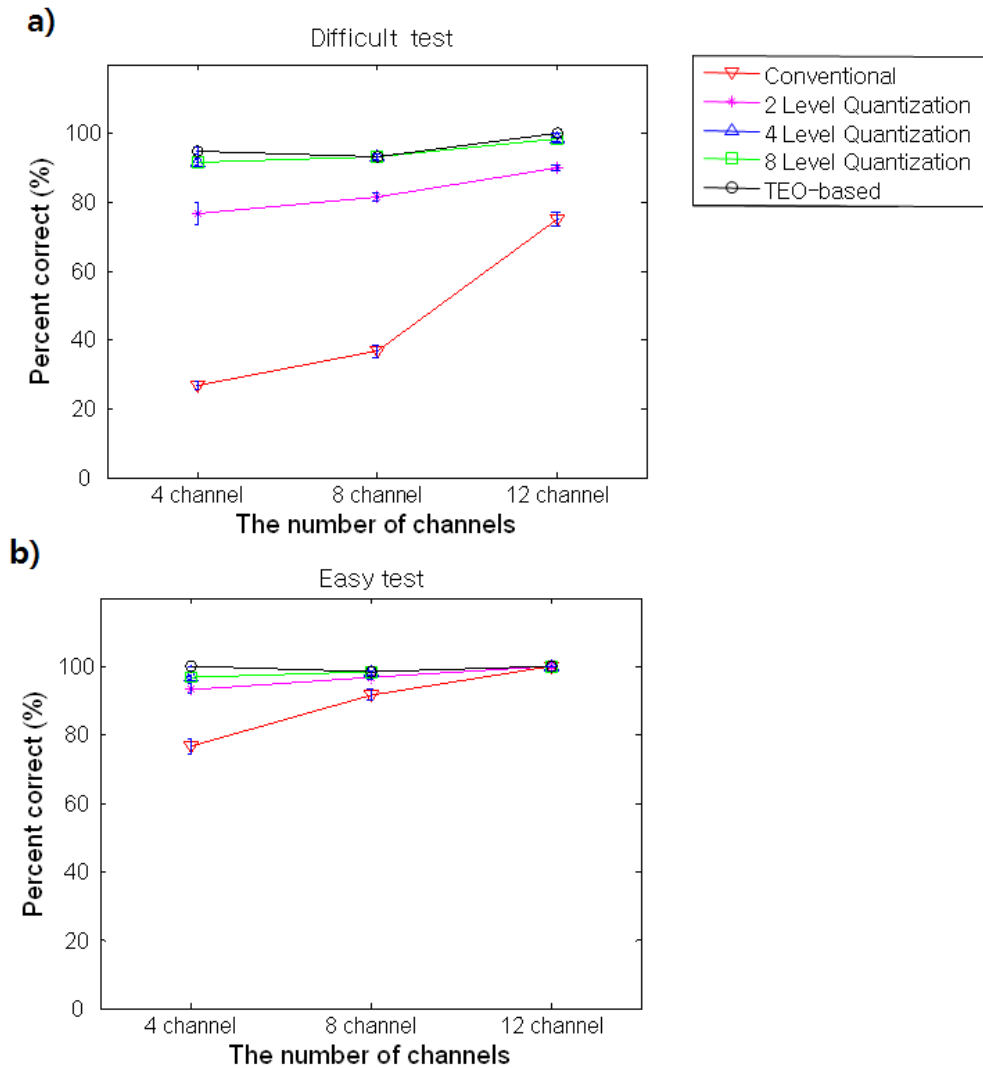


그림 3.6 양자화된 순간주파수로 인한 멜로디인지율 a)난이도 ‘어려움’의 결과 b)난이도 ‘쉬움’의 결과

3.2.2. 양자화된 순간주파수의 영향

그림 3.6은 양자화된 순간주파수가 어음처리기의 멜로디인지율에 영향을 주는 지 알아보기 위하여 실시한 실험의 결과이다. 채널변화에 따른 멜로디인지율 그래프를 그려보면 양자화된 순간주파수에 의한 멜로디인지율도 포락선만을 이용한 어음처리기 인지율에 비해 크게 증가한다 (그림 3.6 (a)). 또한 4레벨과 8레벨로 양자화한 순간주파수를 이용한 어음처리기의 인지율은 본래의 연속적인 순간주파수를 이용한 결과와 거의 비슷하였다. 표 3.1은 그림 3.6 결과에 대해 one-way ANOVA를 한 결과이다. 표를 보면 양자화된 순간주파수는 TEO로 추출한 연속적인 순간주파수와 비슷한 정도의 성능을 가지며, 이는 기존 어음처리기의 멜로디인지율에 비해 유의미하게 증가한 정도이다 ($p < 0.001$).

그림 3.7은 양자화된 순간주파수가 어음처리기의 어음인지율에 영향을 주는 지 알아보기 위하여 실시한 실험의 결과이다. 그래프를 보면 멜로디인지와 비슷하게 양자화된 순간주파수를 사용한 어음처리기도 기존 어음처리기에 비해 높은 어음인지율을 보였다. 표 3.2는 그림 3.7 결과에 대해 one-way ANOVA를 한 결과이

표 3.1 양자화된 순간주파수로 인한 멜로디인지율 (그림 3.6)의 p 값

Speech Processor	연속적인 순간주파수					
	Difficult test			Easy test		
	4 ch.	8 ch.	12 ch.	4 ch.	8 ch.	12 ch.
Conventional	0.000	0.000	0.000	0.003	1.000	1.000
2레벨 양자화	1.000	0.261	0.328	1.000	1.000	1.000
4레벨 양자화	1.000	1.000	1.000	1.000	1.000	1.000
8레벨 양자화	1.000	1.000	1.000	1.000	1.000	1.000

다. 4채널 어음처리기에 대해 2레벨로 양자화한 결과를 제외하면 순간주파수의 양자화가 어음처리기의 어음인지율에 미치는 영향은 미미하였다.

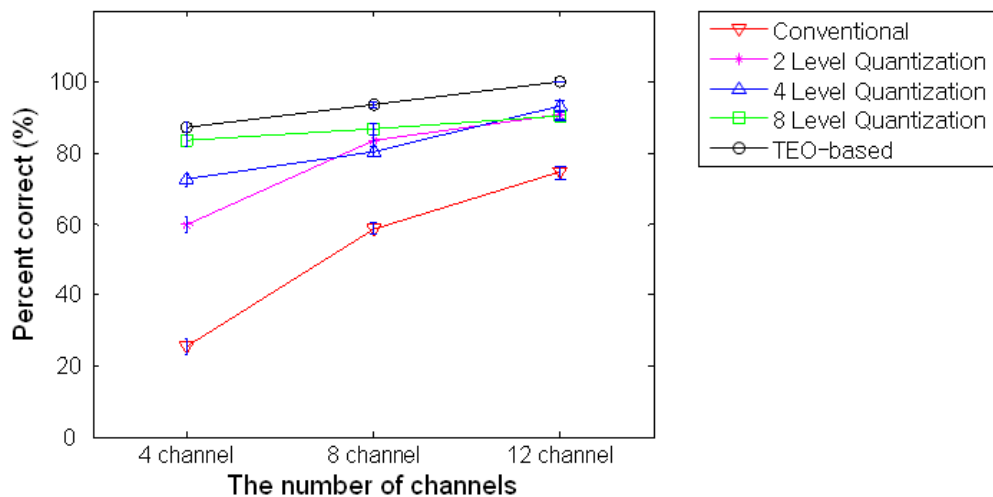


그림 3.7 양자화된 순간주파수로 인한 어음인지율

표 3.2 양자화된 순간주파수로 인한 멜로디인지율 (그림 3.7)의 p 값

Speech Processor	연속적인 순간주파수		
	Difficult test		
	4 ch.	8 ch.	12 ch.
Conventional	0.000	0.000	0.002
2레벨 양자화	0.034	1.000	1.000
4레벨 양자화	1.000	0.450	1.000
8레벨 양자화	1.000	1.000	1.000

3.3 힐버트 변환과 TEO의 연산량 비교

그림 3.8은 힐버트 변환과 TEO의 연산량을 비교하기 위하여 각 방법의 연산 시간을 측정한 결과이다. 실험은 40회 반복하였고, 평균 및 분산을 구하여 그래프를 그렸다. 힐버트 변환에 기반을 둔 16채널 어음처리는 10초길이의 정현파에 대해 평균 15.565초의 연산시간을 가지고, TEO 기반 방법은 9.860초를 가진다. 따라서 TEO 기반 어음처리는 힐버트 변환 기반 어음처리에 비해 36.65% 빠른 연산속도를 가진다. 이 증가폭은 통계적으로 유의미한 값이다 (t-test, $p=2.19 \times 10^{-62}$).

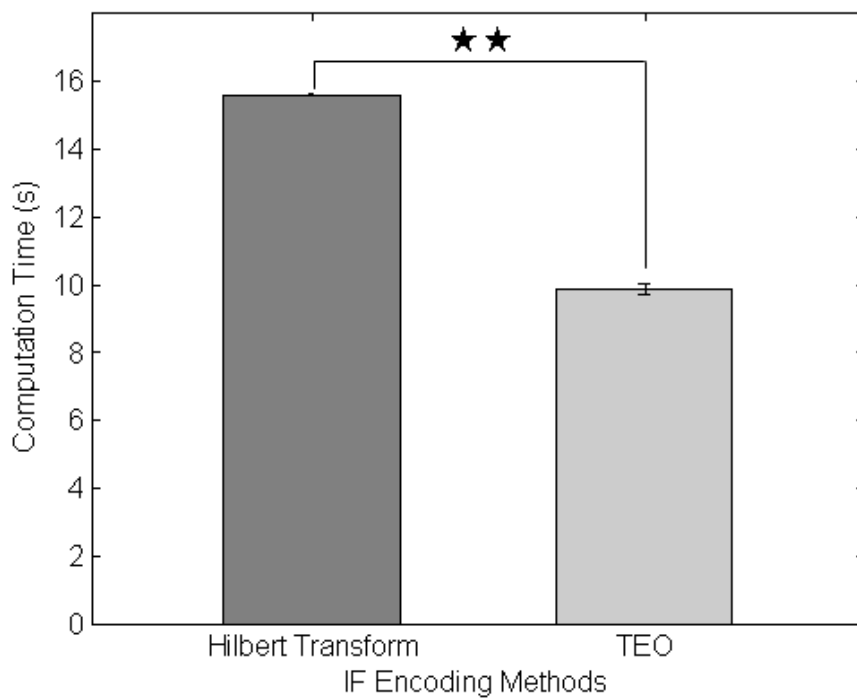


그림 3.8 힐버트 변환과 TEO의 연산속도

제 4 장 고찰 및 결론

본 연구에서는 TEO에 기반을 두어 순간주파수를 추출하고 이를 이용하는 새로운 어음처리를 제안하였다. 또 TEO를 기반으로 하는 방법과 힐버트 변환을 이용한 어음처리가 비슷한 순간주파수 추출성능을 보이지만 TEO를 기반으로 하는 방법이 적은 연산량을 가짐을 밝혔다. 두 방법이 비슷한 멜로디인지율과 어음인지율을 보이는 것을 바탕으로 하여 TEO 기반 순간주파수 추출방법이 인공와우의 어음처리에 적합하다고 결론지었다. 그 이유는 TEO의 적은 연산량이 인공와우의 실시간처리방법에 더욱 유리하며 적은 에너지 소모량을 가지기 때문이다.

포락선 정보만을 가진 어음처리에 비해 순간주파수정보를 추가한 어음처리는 입력음의 주파수정보표현에 탁월한 능력을 가진다. 이는 스펙트로그램 결과인 그림 3.1과 그림 3.2에서 확연히 들어난다. 그림 3.1의 멜로디입력에 대한 두 어음처리의 스펙트로그램을 보면 순간주파수 정보를 추가한 어음처리는 (c,d) 입력음 (a)의 주파수특성을 잘 표현한다. 반면 포락선정보만을 가진 어음처리 (b)는 입력음의 주파수변화를 잘 표현하지 못한다. 특히 F_0 의 변화를 표현하지 못해 출력음이 마치 변화하지 않는 음처럼 들리게 된다. 한편 그림 3.2의 스펙트로그램에서도 순간주파수를 적용한 어음처리의 우수성을 발견할 수 있다. 입력음 (a)의 주파수 특성을 보면 0.2초 부근에서 포먼트가 크게 변화한다. 이는 입력음 /zai/가 /a/에서 /i/로 바뀌는 순간으로 이중모음이라고 일컬어진다. 개선된 어음처리의 스펙트로그램(c,d) 상에서는 이러한 이중모음이 잘 관찰된다. 그러나 포락선 이용 어음처리는 짧은 시간간격에서 일어나는 포먼트변화를 잘 표현하지 못하고 직선으로 표현한다. 한국어단어에서는 이러한 이중모음이 많이 발견되며 이중모음 표현에 강점을 가진 어음처리는 특히 한국인에게 많은 혜택을 줄 것이다 [26].

인공와우의 발전역사에서 음악인지와 관련된 부분이 미약한건 사실이나, 음악은 환자들의 삶의 질을 향상시키는데 중요한 요소이다. 멜로디는 음악의 3가지 요

소인 리듬, 멜로디, 화성중에서 기존 인공와우가 가진 가장 취약한 요소이다 [24, 25]. 게다가 자연의 여러 가지 소리들을 인지하여 더욱더 완벽한 청력회복 실현하기 위해 인공와우환자들의 멜로디인지 향상은 필수적이다. 현재 인공와우 장치들이 통상적으로 사용하고 있는 전극의 수는 16개에서 24개이다. 이러한 전극의 수는 조용한 환경에서의 어음인지에는 만족스러울지 모르겠지만 멜로디 인지를 위해서는 부족한 숫자이다. 그러나 인공와우의 향상된 주파수성분표현을 위해 전극의 수를 무한정 늘릴 수는 없으므로 다른 해결방안이 필요하다. 우리는 순간주파수가 채널수를 늘리는 것의 대안이 될 것이라고 기대하는 바이다. 본 연구에서 실시한 각 어음처리의 멜로디인지실험 (그림 3.3)을 통해 추가된 순간주파수정보가 멜로디인지성능향상에 효과가 있음을 발견할 수 있었다. TEO를 기반으로 한 방법과 힐버트 변환을 이용한 어음처리기 모두 기존 어음처리에 비해 상당히 향상된 멜로디인지율을 보였으며, 두 방법 간의 성능차이는 거의 없었다.

순간주파수를 적용한 어음처리기의 향상된 주파수표현능력은 어음인지에도 도움이 되었다. 단어와 문장을 이용하여 측정한 어음인지율 (그림 3.4, 3.5)이 이를 증명한다. 순간주파수정보를 적용 어음처리기는 어음인지율의 향상을 가져온다. 어음인지의 향상은 잡음이 있을 때 더욱 극명하였다. 이는 잡음에 묻힌 포먼트와 같은 주파수정보를 새로운 어음처리기가 잘 표현하기 때문이다. 포먼트는 모음의 정확한 표현에 필수적인 정보이며 자음의 경우에도 상당히 중요한 정보이다 [27]-[29]. 순간주파수 적용 어음처리기로 인한 어음인지율의 향상은 중국어와 같은 음조언어 (tonal language) 사용자에게 더욱 중요하다. 이 언어들은 소리에 높낮이에 따라 뜻이 변하기 때문에 한국어나 영어와 같은 언어들보다 향상된 주파수표현에 의해 얻는 이득이 더욱 크다 [30].

인공와우의 어음처리에 순간주파수 정보를 추가하여 성능 향상을 시키는 실험은 많이 진행되고 있지만 이는 실제 인공와우에 적용하기 전 시뮬레이션 단계에서의 결과이다. 이를 기존 인공와우에 적용하기 위해서는 청신경 자극펄스를 생성하여야 하는데 펄스생성방법은 아직 정확히 제시되지 않고 있다. 앞서 밝혔듯 힐버트 변환과 TEO를 이용하여 추가한 순간주파수는 변화하는 정도가 너무 빠르다. 저역통과필터를 통해 느리게 변하는 순간주파수를 얻어 사용하여도 어음처리

기 성능에 영향을 주지 않음은 이미 밝혀진 사실이다. 청신경을 자극하는 펄스를 생성하기 위해 시간에 따라 연속적으로 변화하는 주파수를 이용하는 것은 많은 제약이 따른다. 순간주파수를 적용한 어음처리기의 펄스생성방법 개발에 도움이 되기 위하여 본 연구에서는 순간주파수를 양자화하는 방법을 설명하였고 실험하였다. 실험결과 양자화한 순간주파수에 의한 멜로디인지 및 어음인지는 본래의 연속적인 순간주파수의 결과와 유사하였다. 멜로디인지율 (그림 3.6)의 경우 순간주파수를 각각 4레벨과 8레벨로 양자화했을 때의 결과는 TEO로 추출한 연속적인 순간주파수의 결과와 거의 일치하였다. 어음 인지율 (그림 3.7)의 경우에서도 양자화한 순간주파수의 결과가 크게 떨어지지 않아 순간주파수 양자화의 유용성을 입증하였다.

순간주파수 적용 어음처리기에서 순간주파수추출방법을 TEO로 하였을 때의 이점을 밝히기 위해서 연산량측정을 하였으며 기대하였던 대로 TEO의 연산속도가 힐버트 변환보다 크게 앞섰다. 힐버트 변환이 많은 연산량을 가지는 이유는 연산과정 중에 컨볼루션이 속해있기 때문이다. FFT를 이용하는 컨볼루션의 속도가 빠르긴 하지만 TEO는 몇 개의 샘플들만을 이용하여 간단한 연산을 하기 때문에 연산량이 적다. 인공와우의 어음처리기에 사용하는 중앙처리장치인 application-specific intergrated circuits (ASICs)와 digital signal processors (DSPs)의 클럭은 그리 높지 않다. 그 이유는 인공와우는 사용시간을 늘리기 위해 저전력 설계를 하기 때문이다. 이러한 인공와우 환경에서 TEO가 가지는 적은 연산량은 매우 큰 강점이다. 실험결과로 제시한 35%의 연산속도 감소는 MATLAB을 이용하여 측정한 결과이다. MATLAB에서는 전문가들에 의해서 설계된 고성능 FFT가 힐버트변환의 convolution 계산을 위해 사용된다. 반면 TEO에 기반을 둔 순간주파수 추출방법은 아직 완전히 최적화되지 못하였다. 만약 두 방법을 실제 DSP에 적용한다면 연산량 차이는 더욱 커질 것으로 예상된다. 연산의 우세뿐 아니라 TEO가 힐버트 변환과 비슷한 성능을 보이기 때문에 인공와우에 더욱 적합한 방법이라고 할 수 있다.

참 고 문 헌

- [1] B. Wilson, and C. Finley, "Improved speech recognition with cochlear implants," *Nature*, vol. 352, pp. 236-238, 1991.
- [2] P.C. Loizou, "Mimicking the human ear," *Signal Processing Magazine, IEEE*, vol. 15, pp. 101-130, 1998.
- [3] P. Loizou, M. Dorman, and Z. Tu "On the number of channels needed to understand speech," *J. Acoust. Soc. AM.*, vol. 106, pp. 2097-2103, 1999.
- [4] B.S. Wilson, R. Schatzer, E.A. Lopez-Poveda, X.A. Sum, D.T. Lawson, and R.D. Wolford, "Two new directions in speech processor design for cochlear implants," *Ear and Hear.*, vol. 26, pp. 73-81, 2005.
- [5] L. Deng, C. D. Geisler, "A composite auditory model for processing speech sounds," *J. Acoust. Soc. Am.*, vol. 82, pp. 2001-2012, 1987.
- [6] L. Deng, I. Kheirallah, "Dynamic formant tracking of noisy speech using temporal analysis on outputs from a nonlinear cochlear model," *IEEE Trans. Biomed. Eng.*, vol. 40, pp. 456-467, 1993.
- [7] R. Meddis, L. P. O'Mard, E. A. Lopez-Poveda, "A computational algorithm for computing nonlinear auditory frequency selectivity," *J. Acoust. Soc. Am.*, vol. 109, pp. 2852-2861, 2001.
- [8] K. Kim, S. Choi, J. Kim, and D. Kim, "An improved speech processing strategy for cochlear implants based on an active nonlinear filterbank model of the biological cochlea," *IEEE. Trans. Biomed. Eng.*, in press, 2008.
- [9] Z.M. Smith, B. Delgutte, and A.J. Oxenham, "Chimaeric sounds reveal dichotomies in auditory perception," *Nature*, vol. 416, pp. 87-90, 2002.
- [10] S. Brill, A. Moltner, W. Harnisch, J Muller ,and R. Hagen, "Temporal fine structure coding in low frequency channels: speech and prosody understanding, pitch and music perception and subjective benefit evaluated in a prospective

randomized study," *Conference on Implantable Auditory Prostheses*, California, USA, 2007.

[11] W.R. Drennan, J.K. Longnion, C. Ruffin, and J.T. Rubinstein, "Discrimination of Schroeder-phase harmonic complexes by normal-hearing and cochlear-implant listeners," *J. Assoc. Res. Otolaryngol.*, vol. 9, pp. 138-149, 2008.

[12] W.R. Drennan, J.H. Won, V.K. Dasika, and J.T. Rubinstein, "Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise," *J. Assoc. Res. Otolaryngol.*, vol. 8, pp. 373-383, 2007.

[13] F. Zeng, K. Nie, G.S. Stickney, Y. Kong, M. Vongphoe, A. Bhargave, C. Wei, and K. Cao, "Speech recognition with amplitude and frequency modulations," *Proc. Natl. Acad. Sci. USA*, vol. 102, pp. 2293-2298, 2004.

[14] K. Nie, G. Sticney, and F. Zeng, "Encoding frequency modulation to improve cochlear implant performance in noise," *IEEE Trans. Biomed. Eng.*, vol. 52, pp. 64-73, 2004.

[15] R.E. Ziemer, and W.H. Tranter, *Principles of Communications*, Houghton Mifflin, 1990.

[16] J.F. Kaiser, "On a simple algorithm to calculate the energy of a signal," *in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 1, pp. 381-384, 1990.

[17] P. Maragos, J.F. Kaiser, and T.F. Quatieri, "Energy separation in signal modulations with application to speech analysis," *IEEE Trans. Signal Process.*, vol. 41, pp. 3024-3050, 1993.

[18] A. Potamianos, and P. Maragos, "A comparison of the energy operator and the Hilbert transform approach to signal and speech demodulation," *Signal Process.*, vol. 37, pp. 95-120, 1994.

[19] C.S. Throckmorton, M. S. Kucukoglu, J. J. Remus, and L. M. Collins, "Acoustic model investigation of a multiple carrier frequency algorithm for

- encoding fine frequency structure: Implication for cochlear implants," *Hear. Res.*, vol. 218, pp. 30-41, 2006.
- [20] J.T. Rubinstein, and C. Tuner, "A novel acoustic simulation of cochlear implant hearing: effect of temporal fine structure," in Proc. 1st International *IEEE EMBS Conference, Neural Engineering*, Mar. 2003, p. 142-145.
- [21] 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, pp. 630-637, 1999.
- [22] D.H. Kim, J.H. Kim, and K.H. Kim, "Performance evaluation of speech onset representation characteristic of cochlear implants speech processor using spike train decoding," *J. Biomed. Eng. Res.*, vol. 28, pp. 694-702, 2007.
- [23] J.H. Galvin, Q. Fu, and G. Nogaki, "Melodic contour identification by cochlear implant listeners," *Ear and Hear.*, vol. 28, pp. 302-310, 2007.
- [24] M.C. Leal, Y.J. Shin, M-L. Laborde, M-N. Calmels, S. Verges, S. Lugardon, S. Andrieu, O. Deguine, and B. Fraysse, "Music perception in adult cochlear implant recipients," *Acta. Otolaryngol.*, vol. 123, pp. 826-835, 2003.
- [25] C.J. Limb, "Cochlear implant-mediated perception of music," *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 12, pp. 337-340, 2006.
- [26] B.K. Yang, "An Acoustical Study of Korean Diphthongs," *MALSORI*, vol. 25, pp. 3-26, 1993.
- [27] S. Bandyopadhyay, E. D. Young, "Discrimination of voiced stop consonants based on auditory nerve discharges," *J. Neurosci.*, vol. 24, pp. 531-541, Jan. 2004.
- [28] E. D. Young, M. B. Sachs, "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers," *J. Acoust. Soc. Am.*, vol. 66, pp. 1381-1403, Nov. 1979.
- [29] A. R. Palmer, I. M. Winter, C. J. Darwin, "The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig

cochlear nerve and primary-like cochlear nucleus neurons," *J. Acoust. Soc. Am.*, vol. 79, pp. 100-113, Jan. 1986.

[30] Y. Kong, and F. Zeng, "Temporal and spectral cues in Mandarin tone recognition," *J. Acoust. Soc. AM.*, vol. 120, pp. 2830-2840, 2006.

Abstract

An improved cochlear implant speech strategy using instantaneous frequency encoding based on Teager energy operator

Choi, Sung Jin

Dept. of Biomedical Engineering

The Graduate School

Yonsei University

This study presents a improved speech processing strategy incorporating instantaneous frequency (IF) encoding for the enhancement of frequency representation performance of cochlear implants. In addition, this study proposes the use of a Teager energy operator (TEO) for the IF extraction from incoming sound. From time-frequency analysis using spectrograms, we verified that the TEO-based method provides excellent frequency representation performance and the superiority was also found when using Hilbert transform (HT) for the IF extraction. Through acoustic simulation, the melody and speech recognition performance of the proposed speech processing strategy was compared with those of a conventional strategy using envelope information only. Insignificant difference in recognition rate was observed between the TEO-based and HT-based IF encodings, and both showed better performance to the conventional strategy. However, the TEO-based strategy was approximately 35% faster than the HT-based strategy.

We verified that the use of quantized IF has similar performance to the use of original continuous IF. From the acoustic simulation, we compared the melody and speech recognition performance of quantized IF encoding with those of normal IF encoding. As a result, we confirmed that the quantized IF encoding, more than two level, does not decrease the recognition rate of the proposed speech processor.

Key words : Cochlear implant, fine structure, instantaneous frequency, the quantization of instantaneous frequency, acoustic simulation