

12. Articulations in Speech and Sound

Perry Cook

12.1 Introduction

- 언어 발음 지각(perception of speech articulation)¹⁾에의 영향 : 혀, 턱, 입술, 연구개(velum), 후두의 길이(larynx height 등)
- 운동 근육 프로그램(motor program)의 실행을 통한 친숙한 단어들 : 자의적, 자동적으로, 호출되거나 재생 가능하도록 뇌에 저장된 훈련된 근육 제어 신호에 의해 발생 → 대화 인식 체계의 부분이 motor program을 사용 → 대화 소리를 발생시키는 동작들을 인식하는 보기들(templates)로 사용
- 이 장에서 사용되는 음성 기호들

<p>/i/ eee as in beet /I/ ihh as in bit /E/ ehh as in bet /ae/ aaa as in bat /U/ uhh as in but /a/ ahh as in father /r/ r as in rat</p>	<p>/p/ P as in pop /t/ t as in tot /d/ d as in dad /h/ b as in bob /g/ g as in gag /s/ s as in soon /ʃ/ sh as in shoot</p>
---	--

12.2 Spectral Properties of Phonemes²⁾

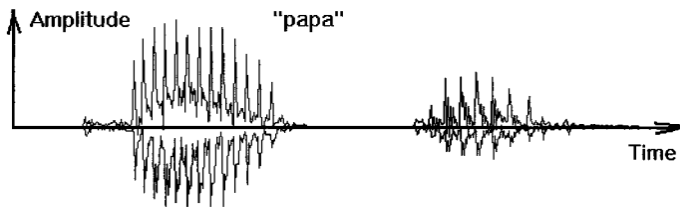


그림 12.1 시간 영역에서의 단어 "papa"의 파형.

- /p/
음성학상 중요, 잡음의 파열(noisy explosions)이후 따라오는 정적이 본질적 특징

1) 네이버 영어 사전 articulation : n. 1 또렷한[명확한] 발음;발음(법);(생각 등의) 명확한 표현 《of》
 2 【음성】 유절(有節) 발음, 개개의 조음(調音);언어(음); 《특히》 자음
 3 【언어】 분절 《발화(發話)의 각 부분을 의미있는 언어음으로 가르기》

2) 네이버 영어 사전 phoneme : n. 【음성】 음소(音素), 음운(音韻) 《어떤 언어에서의 음성상의 최소 단위》

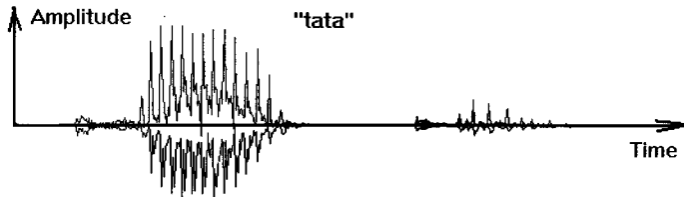


그림 12.2 시간 영역에서의 단어 "tata"의 파형. "papa"와의 유사성(그림 12.1)

- /t/, /p/ 소리

특히 시끄러운 상황이나 전화와 같은 한정된 전달 채널 대역폭에서 흔히 혼동됨

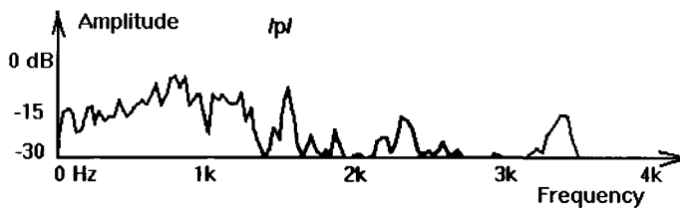


그림 12.4 /p/음의 스펙트럼.

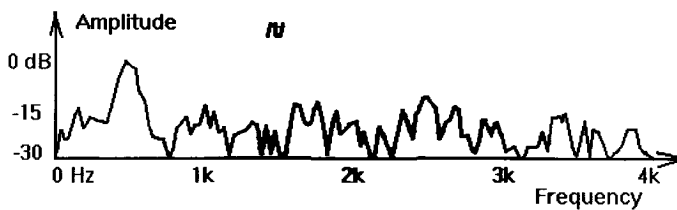


그림 12.5 /t/음의 스펙트럼.

- /p/ 저주파수에 더 많은 에너지
/t/ 4kHz 주파수 영역까지의 모든 영역에 평탄한 에너지

12.3 Production in Perception

- 간단한 정적 음향학(static acoustics)보다 신호들을 사용하여 speech phoneme을 이해
 - 단어와 문장의 맥락
 - 문장의 의미
 - 각 단어의 위치에서 말이 되는 단어들
 - 다른 요소들 : 대화의 인식에 영향을 미칠 수 있는 어떤 특정한 가능한 단어를 말하는 것이 얼마나 어려운지 등
- 초기의 모음과 formant 인지에 대한 실험 결과 : 발생/인지의 관계
- 모음의 발생과 지각에 관한 독창적인 연구(1951)
 - Peterson과 Barney의 결론(1952)
 - a. 발생과 인식은 이전 언어 경험에 달려있음
 - b. 조음상의 기체의 제한된 위치들 때문에 특정한 모음이 다른 모음들보다 더 잘 이해
- 몸짓이 지각적으로 중요한 부분(대화음 자체보다)이라는 개념

- Alvin Liberman의 논문 “On Finding That Speech is Special”(1982)
 - *motor theory of speech perception*
 - 완전히 새로운 음향적 음소 세트의 교육
 - ex. 각 모음을 특징적인 멜로디 악기의 소리로 교체, 각 자음을 타악기 소리로 교체
 - 대화에 사용되는 모음과 자음들과 음향적으로 다름
 - 인간에게는 새로운 세트의 교육 불가(새로운 음소 사전의 단순한 흔한 단어도)
- 외국어를 배우는 경우
 - 친숙하지 않은 소리들
 - 단어 수준과 음소 수준의 표현들의 표현 언어들
 - 대화를 표현하는 대안의 시스템
 - Liberman : 사람들이 자신이 만든 새로운 음향적 음소 사전을 잘 수행하지 못함
 - speech는 모듈적으로(modularly) 특별
 - 뇌에서의 음향 언어 인식 기제 : 본능적으로 발생 기제와 연관, 본질적 분리 불가
 - 발화를 위한 운동 근육 프로그램과 조음 기관에 대한 고유의 한도(limitation)
 - 친^구한 발생 패턴들로 공명하는 복잡한 네트워크를 통해 발화 진행

12.4 Ambiguous Sounds: Ba, Da, Ga

- b, d, g : formant 상각 궤도(trajectory)에서 가장 다름

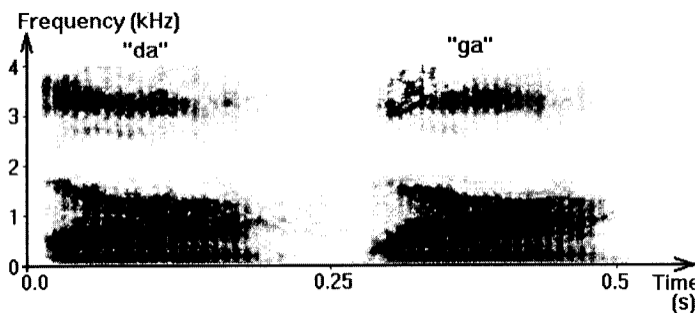


그림 12.5 Spectrograms of the sounds "da" (왼쪽), "ga" (오른쪽) 음의 스펙트로그램. 주요 차이 : 3번째 formant의 상각궤도(the trajectory of the third formant)에 있음.

- masking 현상 : 파열음 이후의 3번째 formant 변화(transition)에 청각적으로 취약
 - “dada”, “daga”, “gada” : 명백히 결정하기 어려움
 - 거꾸로 재생될 경우 transitions : 확실히 다름
 - 모호한 speech sounds : 단순한 즉각적(instant to instant) 음향적 분석보다 더 많은 정보를 사용

12.5 Pick the Simplest Explanation

- “Occam's Razor”(William of Occham, 1285–1349)
 - the simplest explanation is the best.
 - 인간의 지각
 - 3.5장의 예

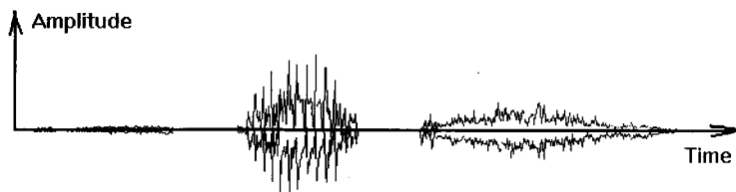


그림 12.6 "sish"를 녹음하여 만들어진 단어의 시간 영역에서의 파형.
"sish"를 녹음하고, 두 부분에 정적을 삽입하면, 대부분은 "stitch"로 인식함.

- 대부분의 사람들 : 전자적으로 편집된 것보다 아는 단어를 사용하는 선호
 - 두뇌 : 음기관들의 연속성을 추정하면서 발화될 것 같은 소리로 하기로 결정
- 연속성의 개념 : 발화의 조음상 묘사(description)와 신체적 움직임의 동작적인 묘사의 중심

12.6 Lexical Filtering and Lazy Evaluation

- /s/, /f/ : 고주파수 영역
 - "ftich" → "speech"로 인식하는 사람들(언어연구에 관련, 주제가 마음에 있음)
- 사전적(lexical³⁾) 접근에 기초한 layer of linguistic filtering
- 여러 개의 비슷한 친숙한 단어에서의 선택 실험
 - "say" 사이에 /s/, /Ei/ → "stay", "spay"
 - : /s/ 발음과 /t/ 발음 형성과 /s/ 발음과 /p/ 발음 형성을 위한 입술의 움직임 시간
 - 완전히 인공적인 방법으로 형성되고 수정되어진 소리의 지각에서도 조음상의 요소가 나타남

3) library of known and common words

12.7 More on Reasonable Formant Ranges

- 신체적 물리적 조음기관들의 제약 : 대화 같은 소리를 내는 적당한 formant 범위와 소리 결정
 - 현실에서 많은 formants들 존재 : 인간이 낸 소리로 지각하지 않음
 - "yah"에서의 /i/(eee), /a/
 - 조음상의 영역에서 interpolation이 존재
 - /i/에서 /a/로 갈 때 가능한 모음 trajectory
- | | | | | | |
|-----|-----|-----|------|-----|-----|
| /i/ | /I/ | /E/ | /ae/ | /U/ | /a/ |
| eee | ihh | ehh | aaa | uhh | ahh |
- 단순한 cross fade나 spectral interpolation과 다름
 - /E/가 중간 조음 지점이 아님 : 6개의 formants가 3500 Hz 이하 지점에 활동

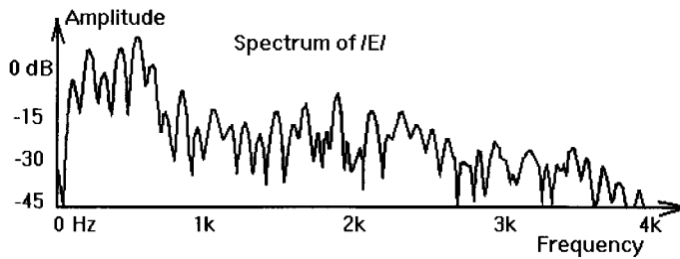


그림 12.7 /ia/의 발음 중 가운데 부분에서의 /E/음의 스펙트럼

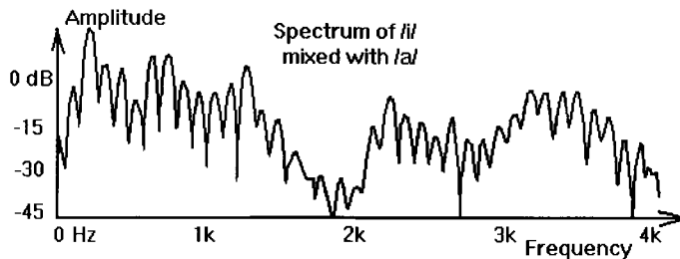


그림 12.8 /a/와 /i/가 섞여 형성된 스펙트럼

- 음성 기제는 조음 속도를 발생시킴 → 발화 중 얼마나 빨리 변화하는가
 - ex. "say"에 정적 삽입 → "stay" "spay"로 인식
 - "stay" : /s/ /t/ 같은 두 개의 혀 수축 지점
 - "spay" : /s/ /p/ 다른 두 개의 혀 수축지점
 - ex. "yah"를 더 빠르게 하여 사람이 조음상의 움직임을 수행할 수 있는 최소 시간 이하로 떨어져 부자연스럽게 들림(조음상의 공간 interpolation이 형성되더라도)
- simple : laziest, easiest, or the utterance that would have taken the least effort.
 - ex. "ara", "ala" : VCV utterances(for vowel-consonant-vowel)
 - "arga", "alda"로 지각 : 조음 위치가 /r/,/g/ 비슷, /l/,/d/ 비슷

12.8 The McGurk Effect⁴⁾, and Prosody and Gesture in Perception

- McGurk Effect(McGurk and MacDonald 1976)의 조음 지각에 대한 실험
 - "ba," "da", "ga"; "bi," "di," "gi" ; 등..
 - 한 가족 중 다른 단어를 말하는 사람 얼굴의 비디오 이미지를 동시에 봄
 - "bi" 오디오와 "gi" 비디오를 동시에 듣고 본 청자 : "di"를 들음
 - "bi" : 입술 끝의 폐쇄로 시작, "gi"는 더 뒤쪽의 혀 윗부분, "di"는 두 조음 지점
중간 지점
- prosody⁵⁾
 - 음정과 음조 곡선의 인지
 - ex. "I wish I could go" 두 번 반복
 - 첫 번째 : "wish" 강조
 - 두 번째 : "I" 강조
 - 단어를 통해 전달되는 정보의 의미에 영향
 - 음율 체계의 지각이 말해지는 단어들에 의해 영향 받는가?
- "gray ship"
 - ex. 선택적으로 정적을 단어에 편집하면 "gray chip", "great chip", "great ship"
청자에게 각 발화의 최대 피치를 예상하라 요구
 - "great ship"은 과대평가
 - "gray chip"은 과소평가
 - ☞ great ship은 어떤 배를 칭송하기 위해 외치는 것 같음 : 높은 피치의 억양을 가짐

12.9 Conclusion

- 발화에서의 조음에 대한 개념
 - concept of gesture : 합성, 분석, 분류화에서 중요
 - picking the simplest, or laziest, or easiest-to-pruduce : 작품(idiom)에서도 보임
- articulation in music performance and perception

4) 스티븐 핀커, 김한영 역, 「언어본능」

http://book.naver.com/bookdb/book_detail.php?where=pvidx_xml&bknu=0306062500&secnum=8&bid=47753&menu=cview&query=

<http://blog.naver.com/yhmpro?Redirect=Log&logNo=80039462247>

http://en.wikipedia.org/wiki/McGurk_effect

5) 네이버 영어 사전 prosody n.

1 작시법;시형론(詩形論), 운율학(韻律學)

2 운율 체계, 시형 Milton's prosody 밀턴의 시형

3 【언어】 운율(론)