

Phonology는 사람이 인식하는 차원이고 Phonetics는 사람이 인식하지 못하는 하위의 차원이다. 발화 시의 물리적인 현상을 그 자체로 바라보는 것이 Phonetics이다.

Vocal tract에는 Nasal과 Oral tract가 있다. Nasal tract로 가면 nasal sound가 되고 oral tract로 가면 oral sound가 된다. Larynx는 후두이고, larynx 사이의 틈을 glottis라고 한다. 사람이 꿀꺽할 때 기도를 막는 역할을 하는게 epiglottis이다. Nasal/Oral은 Velum에 의해 결정되는데, Velum이 raised되면 oral이고 lowered되면 nasal이다. 코로 숨 쉴 때 velum은 lowered된다. Voicing을 결정하는 것은 larynx에서 이루어진다. Glottis가 열려 있으면 성대가 떨리지 않아 voiceless sound가 나오는 반면에 glottis가 닫혀 있으면 성대가 떨려 voiced sound가 나온다.

영어에서 소리는 5가지 요소로 인해 결정이 되는데, Constrictor, Constriction Location(CL), Constriction Degree(CD), Velum, Larynx가 그것이다. Lips, tongue tip, tongue body는 articulatory process를 담당하는 constrictor이다. 이곳에서 혀와 입술을 이용해 소리에 영향을 미친다. Velum은 oro-nasal process를 담당한다. 이곳에서는 소리의 oral/nasal을 결정한다. Larynx에서는 phonation process가 발생하는데, voicing이 이곳에서 결정된다.

Constrictor이 정해졌다면 Constriction location, Constriction degree로 더욱 specify 해야 한다. Lips가 constrictor이라면 bilabial, labiodental이라는 CL로 나뉜다. Tongue tip이 constrictor이라면 dental, alveolar, palato-alveolar, retroflex로 CL이 나뉜다. Tongue body가 constrictor이라면 palatal, velar로 CL이 나뉜다. 'j(y)'는 palatal에 속하고 'sh'는 palate-alveolar에 속하고 'r'은 retroflex에 속한다.

Constriction Degree는 stops, fricatives, approximants, vowels로 나뉜다. Approximants에는 'r', 'l', 'w', 'j' 4개만 있다. 'm', 'n', 'ng'도 nasal sound이지만 stop이다. 모든 모음의 constrictor은 tongue body이다. 모든 phoneme은 이러한 요소들의 combination으로 설명이 된다.

Spectrogram에서 띠로 나타나는 것을 formant라고 한다. 각 띠마다 아래에서부터 first formant, second formant 등으로 부르며 f1, f2 등으로 줄여서 부른다. 아래에서부터 위로 formant는 무한대로 나타난다. Formant값이 모음을 결정한다. 같은 모음이면 비슷한 formant 형태로 나타난다. 즉 formant는 모음을 구별하는 수치적인 지표이다.

Pitch setting- pitch range에서 설정을 해줘야 남자 목소리, 여자 목소리 구별이 가능해진다.

소리를 녹음하고 확대하면 하나의 큰 파도, 패턴이 보이는데, 이 큰 파도 하나에 성대의 떨림 한번이 대응된다. 큰 파도가 1초 동안 나타나는 횟수가 바로 pitch(Hz)이다. 이 큰 파도의 duration을 재고 이 값을 통해 (1 나누기 duration값)을 하면 큰 파도가 1초 동안 나타나는 횟수를 계산할 수 있으며 이것이 pitch값이다. Pure tone으로 내 목소리의 pitch를 입력하면, 내 목소리와 같은 높이의 소리를 들을 수 있다. 이때, pure tone의 곡선은 사인 곡선과 같다.

이 세상 모든 소리는 사인 웨이브의 조합으로 표현된다. 사인 웨이브는 frequency와 amplitude로 결정된다. 사인 웨이브의 x축은 시간이며 y축은 단순한 값(value)이다. 우리 주위의 모든 소리는 pure tone이 아닌 복잡한 소리라고 할 수 있다. 이 때 이 복잡한 소리의 spectral analysis에서 나오는 첫번째 사인 웨이브를  $f_0$ , fundamental frequency, the number of vocal cords vibration of a second 등으로 부른다.  $F_0$ 의 frequency가 전체 소리의 pitch이다.

Source는 성대에서 나는 소리를 그대로 녹음한 소리이다. 성대에서 소리를 바로 녹음하는 것을 EGG라고 한다. Source의 소리와 기타의 소리가 유사하다. Source의 spectrum은 pulse train의 형태를 띈다. 여기서 Pulse는 waveform에서 툭 튀어나오는 부분들을 가리킨다. Pulse train의 spectral analysis를 해보면 gradually decreasing하는 형태를 띈다. Source는 harmonics를 이룬다. Harmonics  $\rightarrow$   $F_0$ 의 배음들( $f_0$ 의 frequency의 배수들)이 이루는 사인 웨이브로 소리가 표현된다. 모든 소리가 harmonics를 이루는 것은 아니다. Source와 기타소리 등이 harmonics를 이루며 gradually decreasing을 한다.

Source는 vocal tract를 거치며 filtered된다. Source가 vocal tract를 지나 articulate되면 peak/mountain과 valley가 있는 산맥 형태의 spectrum이 만들어진다. 즉, 산맥을 만들어 주는 것이 filter의 역할이다. 각 모음을 발음할 때마다의 입, 혀 모양이 다른데, 각 모음마다 좋아하는 산맥형태가 있다고 생각하면 된다. Source가 vocal tract를 지날 때, 그 입모양이 도장 역할을 해서 source의 spectrum을 도장 찍는다고 생각하면 된다. 그렇게 해서 도장 찍힌 spectrum이 실제 발화된 소리의 spectrum이 된다. 이 때 이 spectrum은 산맥을 형성하게 된다. 그 때, 첫 mountain 부분이 첫 번째 formant로  $f_1$ 이며 다음 formant가  $f_2$ 이고,  $f_1$ 과  $f_2$ 만 있으면 모음이 결정된다.  $F_1$ 을 y축,  $f_2$ 를 x축으로 했을 때 입 모양과 똑같이 된다. 이 때  $f_1$ 은 혀의 높낮이를 결정하고,  $f_2$ 는 혀의 front, back를 결정한다.

Spectrogram은 spectrum을 시간축으로 늘어놓은 것이다. Spectrum의 입체형이 spectrogram이라고 생각하면 된다. Spectrogram의 x축은 시간이고 y축은 frequency이고 짙은 정도는 amplitude이다. Spectrum들을 연달아 연속적으로 연결한 것이 spectrogram이다.

Spectral slice를 통해 spectral analysis를 정확히 보기 위해서는 최소 여러 개의 큰 파도(큰 파도 한 번에 성대 한 번 진동)를 선택해야 spectral analysis를 제대로 알 수 있다.

Stereo는 병렬적으로 일일이 다 더해서 나오는 소리이다. Pure tone이 남아있는 소리이다. Mono는 pure tone이 남지 않고 완전히 합쳐지는 소리이다. 그래서 단 하나의 wave form으로 표현되며, spectral analysis를 해보면 mono 소리를 구성하는 각 사인웨이브를 알 수 있다. Mono로 convert를 하면 mono 소리가 그  $f_0$  소리와 pitch가 동일하게 된다. Stereo는 그렇지 않다. Gradually decreasing 하고 harmonics를 이루는 pure tone을 여럿 mono로 만들면 source 소리가 나와 pulse가 만들어지는데, pure tone이 많을수록 pulse가 더욱 두드러지게 튀어나온다.