



Guidance to the Praat, a Software for Speech and Acoustic Analysis

Cheol Jae Seong

Department of Linguistics, Chungnam National University, Daejeon, Korea

음성 및 음향분석 프로그램 Praat의 임상적 활용법

성철재

충남대학교 언어학과

Praat is a useful analysis tool for linguists, engineers, doctors, speech-language pathologists, music majors, and natural scientists. Basic parameters including duration, pitch, energy and perturbation parameters such as jitter and shimmer can be easily measured and manipulated in the sound editor. When a more in-depth analysis is needed, it is recommended to understand the advanced menus of the object window and learn how to use them. Among the object window menus, vowel formant analysis, spectrum analysis, and cepstrum analysis can be cited as useful ones in the clinical field. The spectrum object can be usefully used for voice quality measurement and diagnosis of patients with voice disorders by showing the energy distribution according to frequency axis (domain). A cepstrum object is useful for speech analysis when periodicity of the sound object is not measurable. The low to high ratio obtained from the spectral object and the CPPs measured from the cepstrum object have attracted many researchers, and it has been proven that the CPPs measured in Praat are relatively excellent.

Keywords Praat; Perturbation; Spectrum; Cepstrum.

Received July 5, 2022
Revised August 7, 2022
Accepted August 16, 2022

Corresponding Author

Cheol Jae Seong, PhD
 Department of Linguistics,
 Chungnam National University,
 99 Daehak-ro, Yuseong-gu,
 Daejeon 34134, Korea
Tel +82-42-821-6395
Fax +82-42-823-3667
E-mail cjseong49@gmail.com

ORCID iD

Cheol Jae Seong
<https://orcid.org/0000-0003-1111-0216>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

서론

Praat은 네덜란드 암스테르담 대학교의 Boersma와 Weenink[1]가 1991년도에 처음 공개하였으며, 현재는 동료인 David Weenink와 함께 공동 관리하면서 꾸준히 업데이트를 하고 있는 음성 및 음향분석 소프트웨어다. IBM 호환기종이나 애플, 리눅스, 크롬북, 라스프베리 파이 등 다양한 운영체제에서 실행 가능하다. 웹 사이트를 통해서 무료로 내려받아 사용할 수 있으며, Free Software Foundation의 GNU 일반 공용 라이선스를 따른다. 즉 무료로 배포되지만 상업적 목적으로는 사용할 수 없는 소프트웨어라는 의미다.

Praat은 음성분석과 음향분석 목적으로 사용할 수 있으며 포먼트(formant) 음성합성과 조음 음성합성도 가능하다. 또한 experimentMFC라는 Praat이 제공하는 지각(perception) 실험용 스크립트 프로토콜을 익히면 지각, 인지 실험도 가능하다. 또한 말소리

중재(intervention)나 치료(treatment)용 연습 프로그램을 데모 스크립트를 이용하여 구성해볼 수도 있다. 사용자의 목적에 따라 쓰임새를 다양하게 활용해볼 수 있는 유용한 소프트웨어다.

이 글은 임상(臨床醫) 혹은 언어재활사의 임상 현장에서 유용할 수 있는 Praat의 기능적 측면의 기초적 이론과 사용법이다. Praat의 사용법은 Praat 내부에서 얻을 수 있는 매뉴얼을 읽어보면 알 수 있으나 처음 접하는 이에게는 그리 쉬운 일이 아니다. 많은 시간을 들여 시행착오성의 노력을 하면 익혀갈 수 있는 여타의 프로그램과는 달리 음향음성학적, 음성공학적인 지식이 없으면 접근하기가 상당히 힘들다. 이런 점을 감안하여 임상에 직접적으로 도움이 될 수 있는 현장용 매뉴얼을 겸한 내용으로 작성해보겠다. 형식적으로는 사운드 에디터에서 직접 구할 수 있는 pitch와 에너지(intensity)관련 변수들, 객체창(object window)에서 임상적으로 유용한 객체의 종류들과 이와 관련된 query 메뉴를 중심으로 구성될 것이며, 기능적으로는 사운드 객체에 대한 지속시간, pitch, 에너지 측정, 파형의 주기에 따른 pitch와 에너지 변동률 즉 jitter, shimmer 등의 섭동변수(perturbation parameters) 다루기, 모음을 객관적으로 바라볼 수 있는 포먼트 측정 및 이와 관련된 파생 변수 다루기, 음질의 이상 유무를 가능해볼 수 있는 스펙트럼 관련 변수들과 주기성이 전혀 측정되지 않을 정도로 망가진 목소리 특징도 잡아낼 수 있는 캡스트럼 관련 변수들이 다루어질 것이다.

본 론

사운드 에디터에서의 음향변수 측정

Praat의 에디터는 사운드 에디터와 텍스트그리드 에디터 두 종류로 나누어진다. 사운드 에디터에서는 포먼트 분석, pitch 분석, 강도(intensity) 분석을 할 수 있고 jitter나 shimmer 등의 섭동변수 목록을 보여주는 voice report를 볼 수 있다. 시간축 파형의 특정 시간점을 선택하여 사운드 스펙트럼 정보도 볼 수 있다. 포먼트는 성도(vocal tract)의 공명주파수(resonant frequency)로 성대를 막힌 쪽, 입술을 열린 쪽으로 하는 튜브 모델(one closed end pipe)에서의 물리적 공명 특성을 일컫는다. 포먼트 주파수는 성도 공명이 주요하게 작용하는 모음 및 공명자음에 속하는 /ㄴ, ㄹ, ㅁ, ㅇ/ 등의 비음, 설측음과 탄설음으로 실현되는 /ㄹ/ 유음 등을 분석할 때 중요한 변수가 된다.

Pitch와 에너지

Praat의 메뉴는 객체창 위쪽에 배치된 고정 메뉴(fixed menus)

와 분석 객체에 따라 유동적으로 달라지는 객체창 우측의 동적 메뉴(dynamic menus)로 구분할 수 있다. 고정 메뉴의 [open > read from files...]를 눌러서 분석할 사운드 파일을 불러온 뒤 동적 메뉴의 [view & edit]를 누르면 독립된 윈도우에서 사운드 에디터를 볼 수 있다. Pitch 분석을 위해서 pitch 메뉴의 show pitch를 활성화시킨다. 그 아래쪽 pitch setting을 누르고 들어가면 분석 대상에 맞는 분석 범위(range)를 입력할 수 있다(Fig. 1). Pitch 탐색 알고리즘을 도와주기 위해서 대체로 성인 남성은 75-300 Hz, 성인 여성은 100-500 Hz, 변성기 전까지 어린이들은 120-600 Hz 정도로 세팅한다. 분석 방법은 자동상관(autocorrelation)과 교차상관(cross-correlation) 알고리즘 중에 선택하게 되어 있는데, 문장 운율 분석과 같이 pitch 변동에 민감한 연구주제를 선택했다면 자동상관을 선택하면 좋고, 음질 분석과 같이 성대의 빠른 진동을 연구 대상에 포함시켰다면 교차상관을 고르는 것이 현명하다.

디지털화된 음성신호는 일정한 크기의 분석창(window)을 씌워 일정간격(step)으로 이동하면서 분석하게 된다. 자동상관의 경우 분석창(analysis window)의 크기는 숫자 3을 분석 범위 바닥값으로 나눈 결과값으로 정의된다(3/pitch_floor). 예를들어 성인 남성의 pitch를 분석하기 위해 pitch 바닥값을 75 Hz로 설정했다면 디지털 신호, 시간 영역(time-domain)의 분석창 크기는 0.04 (3/75)초가 된다. 교차상관은 분자값을 1로 정의하므로 분석창 크기가 1/pitch_floor가 된다. 성인 남성의 바닥값을 대입하면 0.013 (1/75)초가 되어 자동상관 알고리즘에 비해 분석창 크기가 작아진다. 성대의 진동과 같은 시간적 변이가 빠르고 주기가 짧은 음향 이벤트를 다룰 때는 분석창 크기가 작은 교차상관이 유리하다. 시간 분해능(time resolution)이 좋기 때문이다. 반대로 창 크기가 커질수록 pitch 분해능이 좋아진다. 시간과 주파수는 서로 반비례하는 관계이기 때문이다.

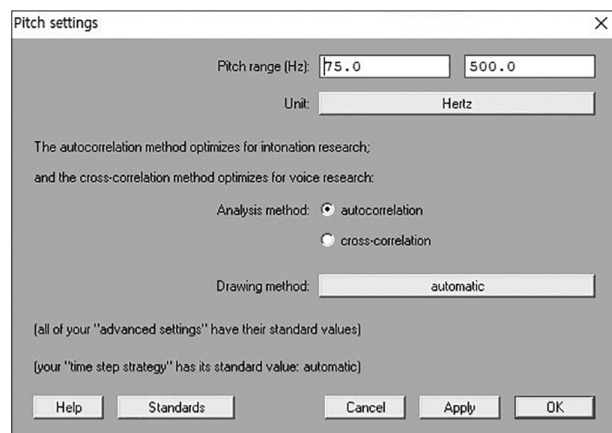


Fig. 1. Dialogue window for pitch setting.

지금까지 설명한 것은 정확한 pitch를 분석하기 위한 pitch 알고리즘 설정에 관한 원칙이었다. 이 원칙을 준수하여 각 데이터의 특성에 맞는 pitch 세팅을 했다면 다음 순서로 에디터 화면에 표시되는 pitch의 그래픽을 적절한 범위로 확장하거나 축소할 수 있게 화면 세팅을 해야한다. 에디터 pitch 메뉴의 서브메뉴인 advanced pitch setting을 이용하면 된다. 성인 남성이 발화한 문장 억양을 분석하기 위해 pitch setting을 이용하여 pitch range 75-300 Hz 범위, 분석방법 autocorrelation

을 선택한 후 대화창을 빠져나가면 Fig. 2와 같이 에디터 오른쪽 쪽 pitch 범위가 75-300으로 표시된다. 왼쪽의 0-5000 Hz 표시는 스펙트로그램의 분석범위를 표시한다. 이 상태에서 pitch의 모양을 조금 더 확대하고 싶으면 advanced pitch setting 창을 이용한다. 이 메뉴는 분석 알고리즘에는 관여하지 않고 보여지는 범위 즉 view range에만 관여하므로 연구자의 필요에 따라 적절한 범위를 입력하기만 하면 된다.

예컨대 에디터 하단 스펙트로그램창에 푸른선으로 나타나

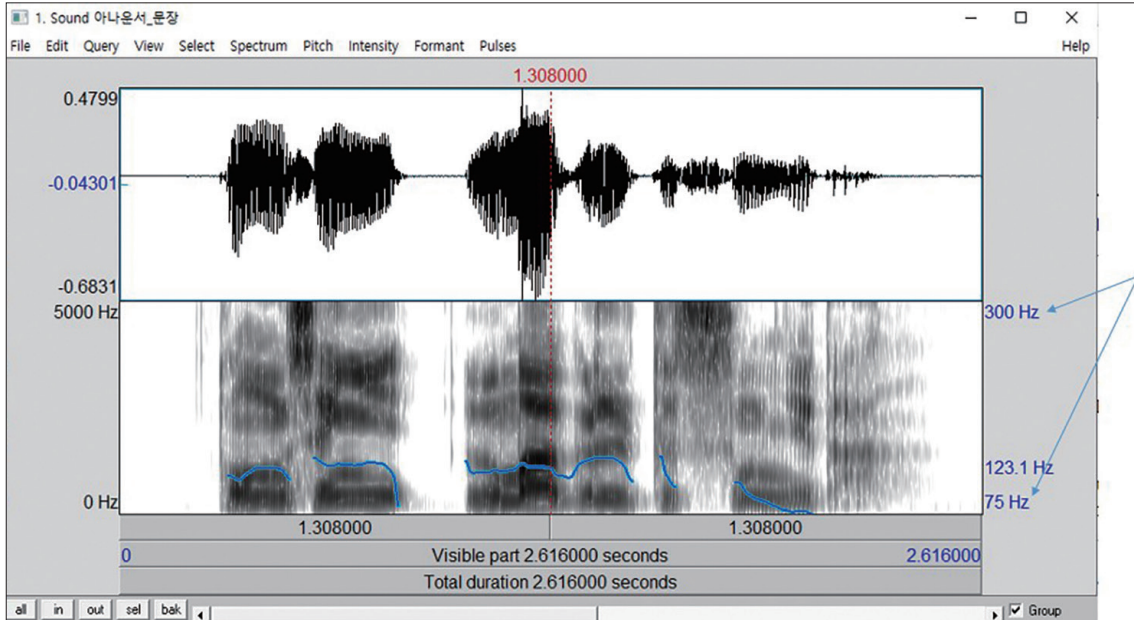


Fig. 2. Screen copy of Sound Editor showing pitch setting ranged from 75 to 300 Hz for adult male.

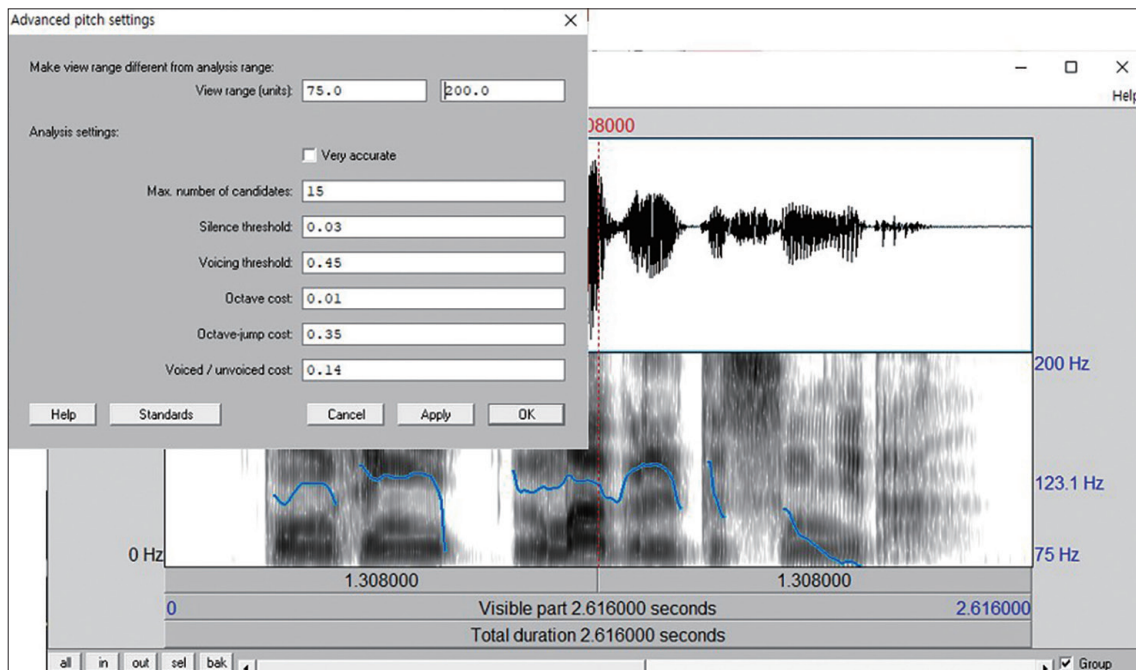


Fig. 3. Screen copy of Sound Editor showing dialogue window for Advanced Pitch Setting ranged from 75 to 300 Hz for adult male.

는 pitch 꺾적선을 좀 더 화면 중앙부위로 이동시키고 상대적 크기를 키우고 싶다면 advanced pitch setting 대화상자의 view range 입력창에 적절한 값(예를 들어 75-200)을 입력해 보면 된다(Fig. 3). Get pitch (단축키 F5), minimum & maximum pitch와 같은 pitch 메뉴의 하위 메뉴를 이용하면 좀 더 편리하게 이용할 수 있다.

사운드 에디터 intensity 메뉴의 하위 메뉴 show intensity 를 누르면 화면에 노란색으로 에너지 꺾적이 활성화된다. 최초(default) 분석 세팅은 50-100 dB 범위로 설정되어 있다(Fig. 4). 즉 이 디폴트 세팅에서는 에디터 화면에 50 dB 이하의 값은 나타나지 않는다. 배경 소음(background noise)을 모두 담고 싶으면 에너지 세팅 범위를 0-100 dB로 설정하면 된다. 조용한 공간에서 녹음하더라도 일정수준의 배경 소음이 상존하고 있으며, 이는 에너지 값 측정에 영향을 미치기 마련이다. 이 소음의 간섭을 근원적으로 제거하기 위하여 Praat은 세팅 설정용 대화상자에 subtract mean pressure를 디폴트 값 on으로 체크해 놓았다. Praat에서 측정하는 intensity 값은 dBSPL로 측정된다. dBSPL은 20 micro (2×10^{-5}) Pascal을 기준치(reference)로 하여 측정하는 값이다. 이 값은 1000 Hz 사인파 조건에서 청각 임계치 규준이 된다[2].

포먼트 측정

정확한 포먼트 측정을 하기 위해서 관련 세팅을 적절하게 해야 한다. 세팅과 관련된 자율권을 사용자에게 주는 부분은 Praat의 최대 강점에 해당한다. 대개의 음성분석 프로그램은

이런 부분이 결여 내지는 충분하지 않다. Praat 매뉴얼에 따르면 성인 남성의 경우 최대 포먼트를 5000 Hz, 포먼트 개수를 5개로, 성인 여성의 경우는 5500 Hz에 5개를 권고하고 있다(Fig. 5)[1]. 이러한 세팅값 차이는 여성과 남성의 성도 크기, 성도 부피의 차를 고려해서 나온 결과일 것이다.

Escudero 등[3]은 Praat을 이용하여 포먼트를 측정할 때 최대 포먼트 주파수와 포먼트 개수 세팅의 다양한 조합을 이용한 측정치 중, 측정구간에서의 포먼트 변이가 최소일 때 가장 적절한 최적의 값이라는 견해를 내놓았다. Yoon과 Kang[4]은 성인 한국어 단모음 데이터를 대상으로 Escudero 등[3]의 포먼트 값 최소변이 세팅으로 검출된 값이 사람의 수동 측정과 높은 양(positive)의 상관관계가 있음을 기술했다. 한국어 포먼트 측정과 관련하여 후설 원순 고모음 /오, 우/와 전설고모음 /이/는 Praat이 권고하는 표준 세팅을 준수했을 때 제 2

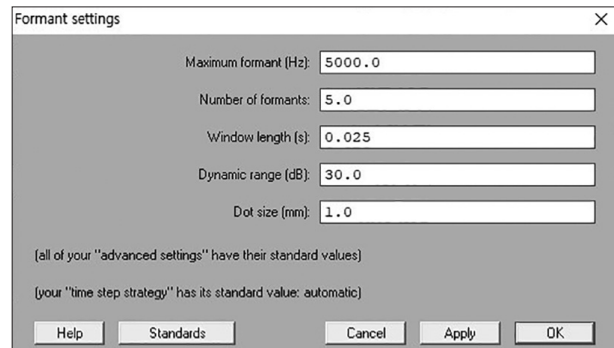


Fig. 5. Formant setting window for adult male, as recommended by Praat manual.

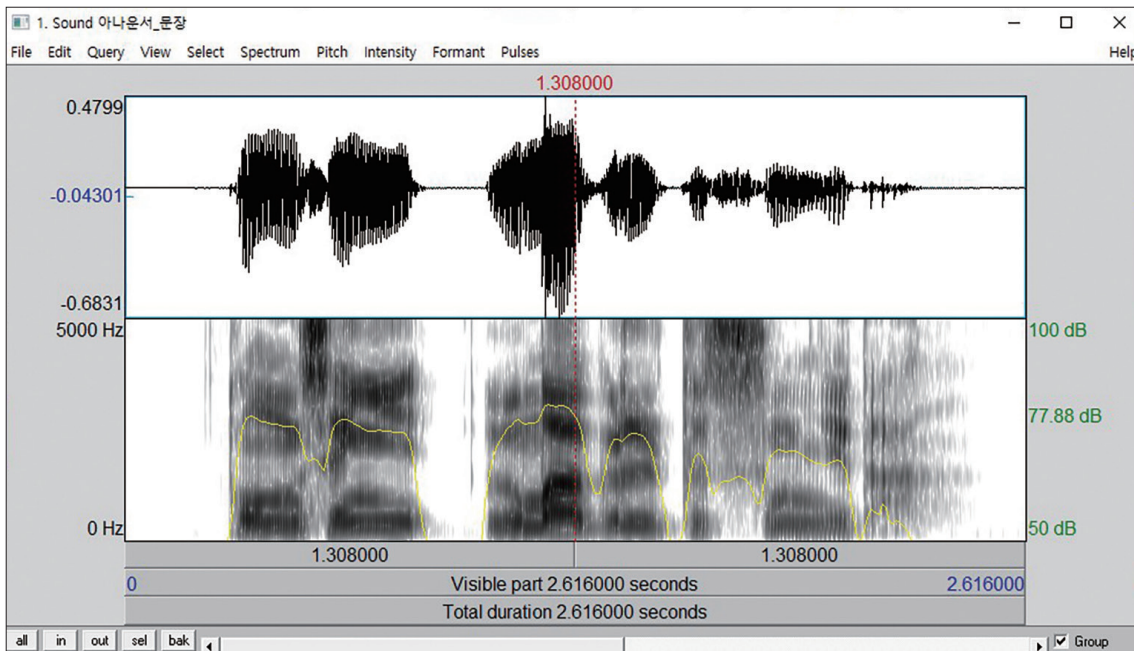


Fig. 4. Screen copy of Sound Editor showing Intensity setting ranged from 50 to 100 dB.

포먼트 값에서 잘못된 측정이 발생할 확률이 꽤 높다[5,6]. 후설 원순 고모음 조음을 위해서 입술이 튀어나오게 되면 전반적으로 성도의 길이가 길어지면서 포먼트 값이 낮아지는데 이를 기존의 디폴트 세팅값으로 보정하기가 어렵다. 전설 고모음은 모음 조음을 위한 조음점(constriction point)이 후치조 경구개 부분에서 이루어지게 되는데 구강 내 F2 수평면의 부피가 극단적으로 줄게 되면서 F2 주파수의 정확한 측정을 어렵게 하는 면이 있다[7].

따라서 Praati 성인 남성용으로 권고하는 최대 포먼트 주파수 5000 Hz, 최대 포먼트 개수 5개와 성인 여성용으로 권하는 최대 포먼트 주파수 5500 Hz, 최대 포먼트 개수 5개는 모음의 종류에 따라 적절히 수정될 필요가 있다. 적절한 세팅값을 찾기 위해서는 최대 포먼트 주파수와 최대 포먼트 개수를 다양하게 조합하여 측정 구간 포먼트 값이 최소의 편차로 측정되는 최적의 조합을 찾을 필요가 있다. 측정하고자 하는 모음의 안정구간(stable section)에서 윈도우 크기와 관련된 디폴트 세팅인 25 ms 윈도우를 이용하여 측정하게 되면 대략 6.25 ms 마다 포먼트 값을 측정하게 된다. 각 윈도우마다 측정된 포먼트 값들의 연쇄가 '최소의 편차'를 보일 때 최적의 포먼트 세팅 조건을 찾았다고 말할 수 있다[8]. 성인 남성이 발화한 /우/ 모음의 안정구간을 대상으로 디폴트 세팅조건(Fig. 5)에서 포먼트를 구해보면 제 1 포먼트(F1)가 329 Hz, 제 2 포먼트(F2)가 652 Hz로 측정되지만, 에디터에서 선택한 구간(Fig. 6, left side)에서 화살표로 표시한 부분과 같이 제 2 포먼트의 흐름에서 편차가 많이 있음을 알 수 있다.

포먼트 세팅을 Park과 Seong[8]의 연구와 같이 4200 Hz/5.5개로 변경해보면 F1 값은 311 Hz, F2 값은 639 Hz로 측정되며 F2의 편차가 상대적으로 줄어들었음을 알 수 있다(Fig. 6, right side). Table 1은 F1과 F2의 편차를 최대로 줄일 수 있는 조건을 여러 세팅을 대상으로 for-loop 반복 스크립트를 돌려 찾아낸 값이다. 한정된 데이터를 대상으로 했으므로 데이터 양이 많아지면 세팅값은 변경될 수 있다. Qtone 컬럼은 적절한 세팅을 찾는 간격을 로그 스케일인 quarter tone 단위로 설정해서 구한 세팅값이다. 아동의 경우 성인에 비해 공명기인

Table 1. Optimal formant settings (along with quarter tone [1/2 semi tone] setting), for each vowel with respect to the sexual difference

Sex	Vowels	Maximum formant (Hz)		Number of formants	
		Linear	Qtone	Linear	Qtone
Adult male	에/이	4700	4627	4.5	4.5
	아/으/어	4300	4303	5.5	5.5
	오/우	4200	4365	5.5	5.5
Adult female	에/이	5400	5599	4.5	4.5
	아/으/어	5300	5205	5	5
	오/우	4400	4911	5.5	5.5
Children (before adolescence)	이	9350	9302	6.7	6.7
	에(애)	9515	9339	6.3	6.3
	아	7380	7156	6.1	6.0
	어	7520	6936	6.3	5.9
	오	5030	5318	6.1	6.0
	우	5590	5911	5.8	6.0
	으	7240	6864	6.0	6.0

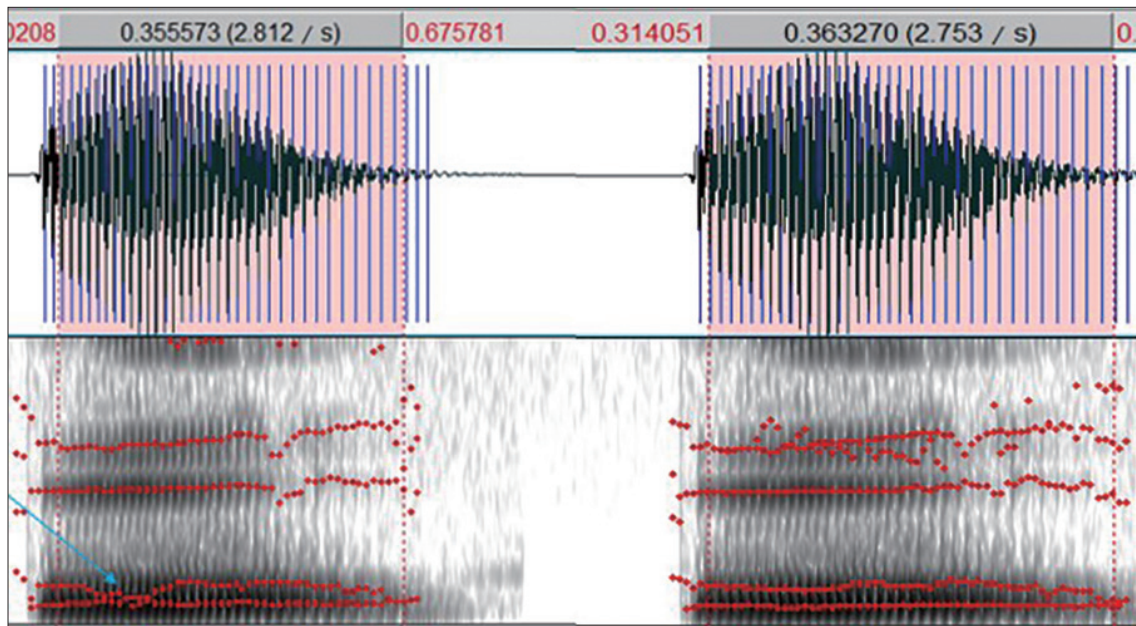


Fig. 6. Formant outputs resulted from two different setting conditions for an [u] from adult male (left: default setting [5000 Hz/5 formants], right: optimal setting [4200 Hz/5.5 formants]).

Table 2. Formant values calculated by optimal script for seven Korean vowels for adult male, adult female, and children

Vowels formant	Adult male (n=16)	Adult female (n=28)	Children (n=20)
이			
F1 (Hz)	279±28	354±55	400±81
F2 (Hz)	2294±148	2965±156	1637±253
에(애)			
F1 (Hz)	455±47	593±61	717±99
F2 (Hz)	1987±138	2473±151	2637±156
아			
F1 (Hz)	740±60	926±65	979±189
F2 (Hz)	1159±131	1385±117	1637±253
어			
F1 (Hz)	548±37	706±62	740±101
F2 (Hz)	905±42	1003±77	1263±118
오			
F1 (Hz)	370±36	438±26	440±71
F2 (Hz)	698±95	774±96	793±156
우			
F1 (Hz)	330±30	417±40	459±63
F2 (Hz)	814±67	912±92	978±209
으			
F1 (Hz)	362±37	443±38	520±56
F2 (Hz)	1356±177	1613±158	1766±147

Data are presented as mean±standard deviation.

성도가 짧고 구강과 비강 또한 부피와 크기가 작기 때문에 성인과는 다른 세팅을 적용해야 올바른 포먼트 값을 측정할 수 있다[9]. Praat 매뉴얼에는 아동용 세팅으로 5500-8000 Hz 정도를 적용해보라는 애매한 문구만 있어서 실제 연구에 적극적으로 활용하기는 어렵다. Table 2는 성인과 아동의 데이터를 대상으로 최적의 포먼트 세팅 알고리즘을 적용하여 구한 F1, F2 값이다. 성인의 경우 기존 권고 세팅을 적용했을 때의 각 포먼트 range (minimum to maximum) 값에 비해 최적 알고리즘을 적용한 경우의 range가 현저히 줄어들었다[8,9]. 예컨대 성인 여성 /이/의 경우 Praat 권고 세팅인 5500 Hz/5개를 적용하면 F2 range 1355 Hz, 최적 세팅인 Linear 스크립트를 적용하면 746 Hz, Qtone 스크립트를 적용하면 747 Hz로 측정되었으며, 아동의 경우 8000 Hz/5개를 적용하면 /이/ F2 range 2814 Hz, 최적 세팅 중 Linear 스크립트에서는 710 Hz, Qtone 스크립트에서는 704 Hz 범위로 조사되었다.

최적의 세팅에 의해 모음 포먼트가 제대로 구해졌다면 이 값들을 이용하여 모음 공간의 특징들을 파악할 수 있다. 모음간 거리는 유클리드 거리값을 이용하여 가늠해볼 수 있으며 모음간 유클리드 거리가 이미 측정되어 있다면 헤론의 공식을 이용하여 구성모음(corner vowels) /이, 아, 우/로 이루어

Table 3. Euclidean distance and Heron's formula applied to 3 corner vowels [i, aa, uu]

Euclidean distance

$$a = \sqrt{(F1_u - F1_i)^2 + (F2_u - F2_i)^2}$$

$$b = \sqrt{(F1_a - F1_i)^2 + (F2_a - F2_i)^2}$$

$$c = \sqrt{(F1_u - F1_a)^2 + (F2_u - F2_a)^2}$$

Heron's formula

$$p = \frac{(a + b + c)}{2}$$

$$area_3(\text{Heron's Formula}) = \sqrt{p(p-a)(p-b)(p-c)}$$

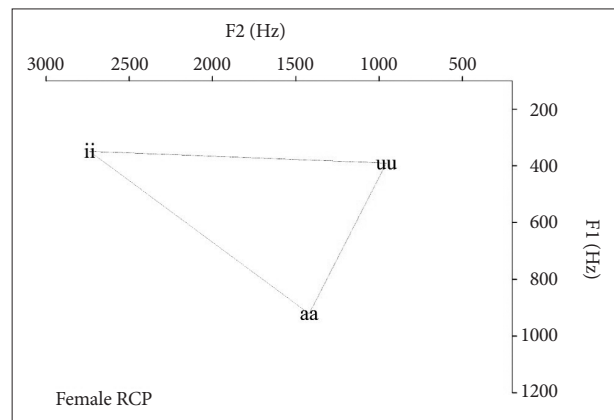


Fig. 7. Vowel space comprised of three corner vowels [i-a-u].

진 모음 삼각형의 면적, /이, 에, 아, 우/로 이루어진 사각형의 면적 등을 구하여 목적에 맞게 이용할 수 있다. 헤론의 공식은 각 변의 길이를 알때 삼각형 면적을 구할 수 있게 하는 공식이므로 사각형이라면 보조선을 그어 두개의 삼각형을 만들고 각 삼각형의 면적을 구하여 합산하면 된다(Table 3).

모음들의 중앙화 경향과 관련하여 몇가지 비율값들이 자주 사용된다. 주로 /이, 아, 우/ 모음 포먼트 값을 이용하여 계산한다. Formant centralization ratio (FCR)은 /이/의 F2와 /아/의 F1을 더한 값을 분모로 하고 /우, 아/의 F2, /이, 우/의 F1을 더한 값을 분자로 하여 계산한다. 분모값이 커질수록 모음 공간이 확장되는 것이니 FCR 값이 작아질수록 확장, 커질수록 축소된다고 이해하면 된다. Vowel articulation index는 FCR의 역수로 정의된다. F2 ratio는 /이/의 F2를 /우/의 F2로 나눈 값이다. 분자가 커질수록 모음 공간의 확장, 분모가 커질수록 모음 공간의 축소로 이해된다(Fig. 7)[10-12].

음질(voice quality) 측정: 섭동변수(perturbation parameters)

말소리의 음질은 음원에 해당하는 성대 진동 뿐만 아니라

여과기에 해당하는 구강성도와 비강성도에 의해서도 좌우된다. 그러나 결정적인 역할은 성대에서 담당한다고 말할 수 있다. Jitter, shimmer와 같은 섭동변수는 성대 진동의 주기성(periodicity)을 기반으로 하여 pitch와 진폭(amplitude)이 얼마나 불규칙적으로 측정 되는가를 알려주는 대표적인 음질 측정 음향변수다. 섭동변수를 측정하기 위해서는 대략 3초 정도의 모음 연장발성(sustained vowel)을 이용해야 한다. 자음과 휴지(pause)가 섞여 있는 문장을 대상으로 하면 안된다. 주기성을 기반으로 한다고 했으므로 음질이 나쁜 정도가 주기성을 측정하지 못할 정도는 아니어야 한다는 전제조건이 있어야 한다. Praat과 같은 프로그램이 의지하는 pitch 측정 알고리즘이 작동할 수 없을 정도로 주기성이 관찰되지 않는다고 한다면 그때는 객체창의 파워스펙트럼 메뉴를 활용하는 캡스트럼 변수를 사용해야 한다. 대표적인 변수로 Cepstral Peak Prominence (CPP)와 Rhmonics to Noise Ratio (RNR)를 들 수 있다.

Praat의 사운드 에디터에서 섭동변수를 측정하기 위해서는 에디터 메뉴 중 가장 오른쪽 pulses에 들어가 show pulses를 활성화시켜야 한다. 그러면 Fig. 8 우측 에디터 화면과 같이 에디터 위쪽 파형창에 푸른색 펄스열이 신호의 주기에 맞추어 생성되어 있음을 확인할 수 있다.

펄스열 생성 후 마우스로 일정 구간을 선택한 후 pulses 메뉴의 하위 메뉴인 voice report를 선택하면 pitch, jitter, shimmer, harmonics to noise ratio (HNR) 등의 섭동변수값을 확인할 수 있는 정보창을 볼 수 있게 된다.

Computerized Speech Lab (CSL, KayPENTAX, Montvale, NJ, USA)의 Multi-Dimensional Voice Program (MDVP)와 Praat은 섭동변수를 측정할 수 있는 이 분야 소프트웨어의 양대 산맥이라 할 수 있다. MDVP는 peak-picking 방식을, Praat은 waveform-matching 방식을 각자의 pitch 측정 알고리즘으로 채택하고 있다. Peak-picking 방식은 각 주기

별 최대 에너지 피크점을 연속적으로 찾아가는 방법을 사용하며, waveform matching 방식은 한 주기와 다음 주기의 waveform 모양이 최대한 일치하는 구간을 연속적으로 찾아가는 방식을 사용하고 있다.

Boersma[13]는 Praat이 채택하고 있는 waveform-matching 알고리즘의 상대적 우수성을 명시적으로 다음과 같이 보고하고 있다: 2초 지속시간의 컴퓨터 합성 모음(기본주파수 117 Hz, F1: 820 Hz, F2: 1300 Hz, F3: 2300 Hz, & 6개의 상위 포먼트로 구성)을 이용하여 인위적으로 1% (underlying)의 local jitter가 생성되게 변조했을 때 waveform-matching 알고리즘과 MDVP의 peak-picking 알고리즘이 구한 local jitter 값에 차이가 없었다. 즉 잡음이 섞이지 않은 병리적이지 않은 음원을 측정 대상으로 할 때 두가지 알고리즘은 서로 차이가 나지 않았다. 그러나 이 합성 모음 시료에 원음의 1%에 해당하는 만큼의 백색잡음(white noise)을 더한 뒤 jitter를 측정해보았을 때 waveform-matching에서는 jitter 0.02% 정도, peak-picking에서는 0.56% 정도로 큰 차이가 발생했다. Waveform-matching 방식은 주기간 큰 차이를 보이지 않았으나 peak-picking 방법은 피크 지점 판단에 대한 알고리즘의 미세한 오류로 인해 주기 추정에 꽤 큰 차이가 났다. 이 주기차를 반영한 jitter 값이 0.56% 정도다. Waveform-matching 방식은 대략 jitter 0.05% 이상이 될 때부터 신뢰할 만하며, peak-picking 방식은 1% 이상 지점부터 믿을 만한 값으로 인정될 수 있다. 이러한 관점을 종합해보면 peak-picking 알고리즘은 병리적 음성이라고 할만한 1.03% jitter 값 이상일 때 신뢰할 수 있을 것이라는 결론을 내릴 수 있으나, Praat의 waveform-matching 방식을 적용한다면 1.03% 이상일 때 병리적 음성이라는 판단은 좀 과한 잣대가 되며, 이보다 낮은 값으로 절사값(cutoff)을 설정해야 한다는 것을 짐작할 수 있다.

Praat의 voice report를 참고하면 기본주파수와 관련하여 선택구간의 pitch 중앙값, 평균값, 최대값, 최소값을 알 수 있

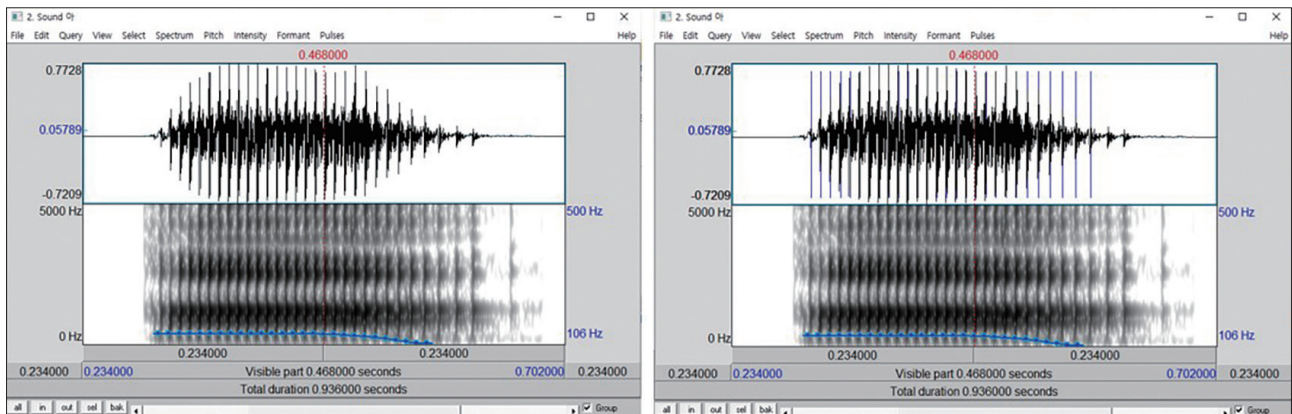


Fig. 8. Sound editor with pulse activated (right top), as opposed to without being activated (left top).

으며 jitter 관련 5가지 변수, shimmer 관련 6가지 변수, harmonicity와 관련하여 noise to harmonics ratio와 HNR 값을 함께 살펴볼 수 있다. Jitter는 주파수 변동률로 정의되므로 주기 간 편차가 어느 정도인지를 가늠해볼 수 있는 음향변수다. Jitter 관련 5가지 변수중 대개의 연구들에서 공통적으로 언급하는 값은 jitter local 값이다. Jitter local과 함께 jitter rap, ppq5, ddp는 단위가 %이며 jitter (local, absolute) 값은 비율값이 아니기 때문에 초(sec) 단위를 사용한다. Jitter (local, absolute) 값은 측정 구간 내 연속하는 두 주기의 절대값 차 평균을 [측정구간 전체 주기수 -1]로 나눈 값이다. Jitter local은 jitter (local, absolute) 값을 평균주기로 나눈 값이며 (jitter=jitter [absolute seconds]/meanPeriod [seconds]) 신뢰할만한 연구들에서 공통적으로 다루고 있는 변수다.

Shimmer는 진폭 변동률로 정의되므로 각 주기간 진폭이 어느 정도 변동이 되는지를 알 수 있게 해주는 변수다. Praat이 제공하는 shimmer 관련 변수는 shimmer (local, %), shimmer (local, dB), shimmer apq3, shimmer apq5, shimmer apq11, shimmer ddp로 구성되어 있다. Shimmer (local, dB)는 연속하는 두 주기의 진폭 간 절대값 차 평균에 상용로그를 취한 뒤 20을 곱한 값으로 단위는 dB이며 나머지 shimmer 값은 %다. 여러 연구들에서 공통적으로 보고하고 있는 대표적 shimmer 값은 shimmer local이며 연속하는 두 주기의 진폭값의 절대값 차 평균을 구간 평균 진폭으로 나눈 값이다.

음질에 문제가 없는 정상 성인의 /아/ 모음 연장발성 데이터를 중심으로 jitter local과 shimmer local 값을 보고하고 있는 국내 연구 중 적절하게 측정하였다고 평가할 만한 몇가지를 소개하면 다음과 같다. Ko 등[14]은 음질에 문제가 없는 정상 성인 남녀 각 40명을 대상으로 /아/ 연장발성 1.5초 구간의 jitter local 값을 남성 0.333 (± 0.081)%, 여성 0.312 (± 0.120)%로 보고했으며, shimmer local은 남성 2.048 (± 0.806)%, 여성 1.857 (± 0.750)%로 보고했으나 아래의 연구보다 값이 작은 경향이 있다.

Seo와 Shin[15]은 남성 132명, 여성 179명을 대상으로 /아/ 연장발성 2.8초 이상 구간의 jitter local 값으로 남성 0.24 (± 0.15)%, 여성 0.14 (± 0.11)%를, shimmer local 값으로 남성 6.05 (± 5.16)%, 여성 5.90 (± 4.69)%를 제시하고 있다. Jeon [16]은 음성피로가 없는 성인 여성 25인을 대상으로 /아/ 연장발성 2초 구간의 jitter local 값 0.341 (± 0.020)%, shimmer local 값 4.958 (± 2.635)%로 측정하였다. Jitter 값은 대체적으로 측정 시료의 구간 길이가 길어지면 값이 작아지는 경향이 있다. 이상의 보고내용을 종합하면 정상 성인 남성의 경우

jitter local 0.414%, shimmer local 11.21%를 초과하면 병리적 음성 권역으로 분류될 수 있는 개연성이 있으며, 성인 여성의 경우는 jitter local 0.361%, shimmer local 10.59%를 초과하면 병리적 음성이라고 잠정적으로 정리할 수 있다. 섭동 변수 측정값은 연구에 따라서 편차가 매우 심하다. 따라서 여기 소개한 값이 절대적인 기준이라고 할 수는 없다.

객체창을 이용한 음향변수 측정

음질 측정: 스펙트럼 변수(spectral parameters)

파열음, 마찰음, 파찰음과 같은 장애음 소음구간의 특징을 파악하기 위하여 적률분석(moment analysis)을 할 수 있다. 파열음 분석을 위한 파라미터로 개방 파열후 후행 모음의 진동이 시작되기까지의 유성개시시간(voice onset time), 후행 모음의 pitch, 후행모음의 H1-H2 (역필터링한 음원 성분 [source]에서 측정한 첫번째 조화음[harmonic]과 두번째 조화음 에너지 차), H1-A2 (역필터링한 source에서 측정한 첫번째 조화음 에너지와 필터링하지 않은 상태의 모음 F2 [제 2 포먼트] 에너지) 등이 보편적으로 사용되어 왔다. 적률분석은 파열개방 구간 스펙트럼 에너지의 분포를 기술통계적으로 파악하여 M1부터 M4까지의 특징변수로 나타낸다[17]. 적률분석에 사용되는 네가지 음향변수들은 마찰음과 파찰음의 소음 구간 분석에서 효율적으로 활용되고 있다.

M1은 에너지의 평균과 같은 개념으로 Praat에서는 무게중심(center of gravity)으로 메뉴화되어 있다. 무게중심은 주파수 축을 양분했을 때 좌/우 에너지 값이 동일해지는 주파수를 말한다. M2는 주파수축(주파수 영역; frequency domain) 에너지들이 평균을 중심으로 얼마나 퍼져 있는지를 알려주는 분산(variance)의 개념으로 Praat에서는 표준편차로 설계되어 있다. M3는 정규분포곡선 즉 가우스 분포(Gaussian distribution)에서 좌, 우 어느쪽으로 데이터가 치우쳐 있는지를 (주파수 영역에서는 좌, 우 어느 쪽으로 에너지가 치우쳐 있나) 살펴보는 왜도(skewness)로 값이 클수록 데이터는 좌측 편포 경향을 보인다. M4는 스펙트럼의 뾰족한 정도(peakedness)를 알려주는 첨도(kurtosis)다. 0을 기준으로 해서 이보다 클수록 뾰족해진다.

Praat에서는 사운드 객체가 유성음이라면 옥타브 당 6 dB 증폭을 위하여 50 Hz 이상 조건에서 고주파 증폭(preemphasis)한 다음 Fast Fourier Transform (FFT)하여 스펙트럼 객체를 생성하고 여기에 연동된 Query 메뉴를 이용하여 구할 수 있다. 한국어 마찰음 /s/은 /이/ 모음이 후행할 때 조음 위치가 치조에서 치조-경구개(alveolo palatal) 위치로 바뀐다[6]. 치조음에 비해서 조음 위치가 입 뒤쪽으로 더 후퇴가 되었으

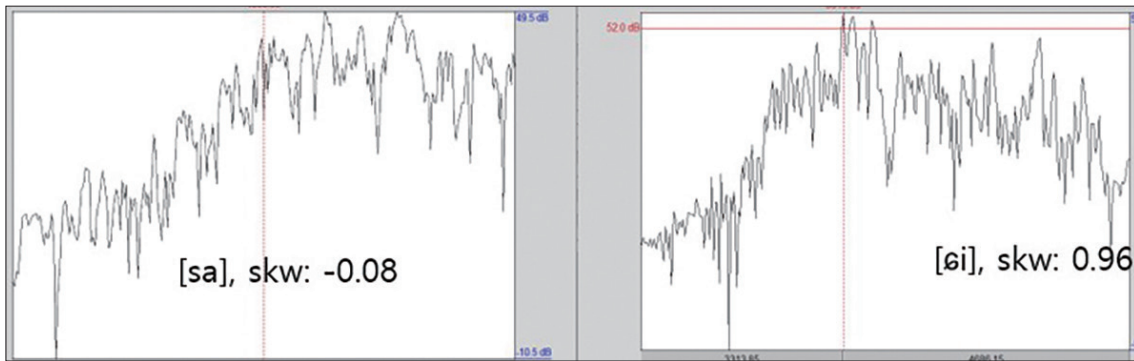


Fig. 9. Comparison of Korean [s] in [sa] and [ʃ] in [ʃi] spectrum, measured in initial 20 ms of frication noise.

므로 [ʃ]는 전강(front cavity)의 부피가 상대적으로 더 크다. 이러한 결과는 {s}의 [s]에 비해 스펙트럴 에너지가 더 저주파수 대역으로 분포하게 됨을 예상하게 한다. 따라서 [ʃ]의 왜도는 [s]에 비해 더 높은 값(positive)으로 측정된다(Fig. 9).

적률분석을 하기 위한 순서를 기술해보면 다음과 같다: 객체창에 사운드 객체를 불러온 다음 사운드 에디터를 열어서 측정구간을 결정하고 필요한 부분을 추출한다. 객체창에서 추출된 부분을 고주파 증폭한다. 우측 동적 메뉴의 filter 탭에 들어가서 filter (pre-emphasis) 명령어(command)를 선택한 후 시작 주파수를 50 Hz 디폴트 세팅 그대로 두고 ok를 클릭하면 고주파 증폭 과정이 끝난다. 이어서 동적메뉴의 analyze 섹션 아래 spectrum 메뉴의 to spectrum...을 선택하고 fast가 디폴트로 체크되어 있는 상태에서 ok를 클릭하면 스펙트럼 객체가 생성된다. Query 메뉴에 들어가 아래쪽에서 무게중심, 표준편차, 왜도, 첨도를 선택하여 값을 구할 수 있다. 이 메뉴들은 가중치 제곱수 p를 요구하는데 디폴트는 2.0으로 입력되어 있으며 그대로 이용하면 된다.

스펙트럼 객체를 이용하여 구할 수 있는 음향변수 중 low to high (LH) ratio가 있다[18]. 사운드 스펙트럼의 0-4000 Hz 저주파 영역 에너지를 4000-8000 Hz 고주파 영역 에너지 값으로 나눈 비율 값이다. 0-4000 Hz까지 주요 포먼트 대역이 몰려있는 영역 에너지(분자)가 약화될수록, 그리고 상대적으로 소음성 잡음이 많이 섞인 4000-8000 Hz 영역 고주파 에너지(분모)가 강해질수록, 즉 LH 비율이 작아질수록 병리적 음성에 가깝다고 해석한다. 이 음향변수는 Praat에서 메뉴를 이용하여 바로 구할 수는 없고 몇가지 절차를 거쳐야 한다. 스펙트럼 객체에 연동된 query 메뉴에서 get band energy... 명령어를 선택한다. 바닥(floor)과 천정(ceiling)값 입력을 요구하는 두개의 필드에 0과 4000을 입력하면 저주파 영역의 에너지를 얻을 수 있다. 동일한 방법으로 4000과 8000을 입력하면 고주파 영역 에너지를 정보창(info window)에 출력할 수 있다. 스펙트럼 객체에서 돌려주는 에너지 값은 파스칼 단위로

출력되는데 엄밀히는 초당 제곱 파스칼로 표시된다(Pa²/sec). 파스칼 단위로 비율을 구하면 비율값이 좀 커지는 경향이 있어서 이를 우리에게 익숙한 dB SPL 단위로 바꾸어 사용할 수도 있다. 변환 공식은 식(1)과 같다. 로그 윗수 분자에 파스칼 단위로 구한 에너지 값을 입력하면 된다.

$$dB SPL = 10 * \log_{10} \frac{pascal^2}{0.00002^2} \quad (1)$$

음성장애가 없는 20대 성인 여성 50명의 /아/ 모음 2초 연장발성을 대상으로 LH 비율을 살펴본 결과 파스칼 단위 비율값은 평균 63843.3 (±283344.8), dB SPL 단위로 변환한 후의 비율값은 평균 1.918 (±0.392)이었다[16]. 성대결절 혹은 성대폴립 진단을 받은 20대 성인 56명(남성 16, 여성 40)의 /아/ 모음 3초 연장발성을 대상으로 LH 비율을 살펴본 결과 파스칼 단위 비율값은 평균 849.08 (±1245.73), dB SPL 단위로 변환한 후의 비율값은 평균 1.525 (±0.139)였다[19]. 관련 선행 연구의 경우, 기능적 음성장애 환자 102명(59.4±12.5세), 정상 집단 59명(52.9±9.1세)의 모음 연장발성 2초를 대상으로 음성장애 환자 699.1 (653.2), 정상 성인 1235.1 (1492.6)을 보고하고 있으며 receiver operating characteristics (ROC)를 이용한 두 집단 절사값으로 369.04를 제시하였다[20].

스펙트럼 객체와 연동된 동적메뉴의 analyse 섹션에서 to LTAS... 혹은 to LTAS (1-to1)...을 선택하면 장구간 평균 스펙트럼(long term average spectrum, LTAS)을 구할 수 있다. LTAS (1-to-1) 객체는 사운드 스펙트럼의 디폴트 bin(주파수 축의 분석 단위) 두께인 1.345825 Hz를 그대로 반영하여 전환되는데 비해 LTAS 객체는 bin 두께를 연구자 마음대로 지정할 수 있다. 디폴트는 1000 Hz로 되어 있는데 연구 목적에 따라 50 Hz 정도로 낮춰서 사용할 수 있다. bin이 두꺼울수록 스펙트럼 에너지 분포의 세부적 특징은 사라지고 대체적 경향성만 남게 된다. 장구간 평균 스펙트럼 객체는 스펙트럼 객체와 달리 밴드 에너지 평균값을 dB SPL 값으로 돌려준다. 스펙트럼 객체에서의 밴드 에너지는 에너지 총량임에 비해

여기서는 평균값임을 주의해야 한다. LH 비율의 고주파 영역인 4000-8000 Hz 밴드 에너지 평균값이(대체적으로) 음수(-)로 출력되므로 비율값을 구하는데 적절하지 못한 단점이 있다. 굳이 LTAS를 이용하여 LH 비율을 구하고자 한다면 query 메뉴에서 4000-8000 Hz 밴드 최소 에너지를 먼저 구한 다음 이 값을 이용하여 전체 에너지를 정규화(normalize) 한 후 계산하면 된다. 정규화는 동적메뉴 modify>formula 섹션에 들어가서 정규화 식을 작성하면 된다. 예컨대 4000-8000 Hz 상위 밴드 에너지 평균이 마이너스로 나왔다면 그 값의 플러스 절댓값을 0-8000 Hz까지 전 영역에 걸쳐 더해지면 되며 수식은 {self+'number'}와 같이 하면 된다.

LTAS 객체의 query 메뉴에는 스펙트럼의 기울기를 계산할 수 있는 두개의 모듈이 있다. Get slope... 명령어의 경우는 저주파대역의 에너지 평균값과 고주파대역 에너지 평균값 차이를 dB로 계산하여 돌려주는데, 디폴트 값으로 설정된 저주파대역은 0-1000 Hz, 고주파대역은 1000-4000 Hz다. 저주파와 고주파대역은 연구자의 목적과 데이터에 따라 설정값을 달리할 수 있다. LH 비율과 마찬가지로 0-4000 Hz, 4000-8000 Hz를 저/고주파수 대역으로 설정하면 음향변수들 사이의 관련성을 좀 더 체계적으로 파악할 수 있을 것이다.

Report spectral tilt... 명령어는 원하는 스펙트럼 주파수 대역의 시작과 끝 주파수를 자유롭게 입력할 수 있게 되어 있고(디폴트는 100-5000 Hz) 주파수 축을 선형과 로그 스케일 중에 선택할 수 있게 되어 있다. 기울기 측정 방식은 고전적 방식인 최소제곱법(least square)과 최근의 알고리즘인 Theil[21]의 Robust 방식 중에 임의로 선택할 수 있다. 결과값에는 y 절편에 해당하는 offset 값과 기울기값(dB/Hz)이 포함된다. 이 명령어는 비교적 최근 버전부터 나오기 시작했으므로 과거 버전에만 익숙한 연구자들은 생소할 수도 있다. 구하고 싶은 스펙트럼 밴드의 기울기를 복잡한 코딩없이 구할 수 있게 되었

으므로 편리하게 되었다. 다만 출력값이 -0.000844와 같이 소수점 아래 매우 작은 값으로 출력되므로 Praat 스크립팅이 가능한 연구자라면 1000배 정도를 곱해서 사용하면 편리하다. LTAS 객체와 연관된 동적메뉴 중 convert 섹션의 compute trend line... 탭을 선택하면 선택한 주파수 범위의 기울기를 그림창에 그릴 수 있다. Fig. 10은 100 Hz bin 조건에서 생성한 LTAS 객체와 tilt를 함께 보여주고 있다.

음질 측정: 캡스트럼 변수(cepstral parameters)

빠른 푸리에 변환(FFT)을 거쳐 얻은 파워스펙트럼에 로그를 취한 뒤 역이산푸리에 변형(inverse discrete fourier transform)을 하면 캡스트럼 객체를 얻을 수 있다. 캡스트럼은 규칙적인 조화음(harmonics) 배열로 설명되는 source 성분과 포먼트 구조로 대표되는 성도에서의 공명 성분을 큐프런시(quefrensy) 축에서 분리하여 보여준다. 큐프런시 축의 좌측 저주파 영역(실상 큐프런시 축은 시간(s)축이지만 주파수 상용 개념이라서 주파수 축의 개념으로 설명한다)에서는 성도의 공명 성분인 포먼트 윤곽(envelope)이 축약되어 들어 있고, 그보다 높은 고주파 영역에서는 조화음 배열이 규칙적으로 나타난다. Praat에서는 일반 사운드 객체를 스펙트럼으로 변형한 뒤 스펙트럼 객체와 연결된 동적메뉴 중 analyze 섹션 아래 to PowerCepstum 명령어로 파워캡스트럼 객체를 생성할 수 있다. 객체가 생성되면 query 메뉴에서 다양한 캡스트럼 변수들을 활용할 수 있다.

캡스트럼 관련 변수들 중 가장 보편적으로 쓰이고 있는 CPP는 Hillenbrand 등[22]이 처음으로 소개하였다. 언급한 섭동변수들은 신호의 주기가 측정될 수 있어야 구할 수 있는 파라미터이지만, CPP는 심각한 음성장애 환자와 같이 주기성을 찾기 힘든 대상의 음성 데이터일지라도 측정이 가능하다는 장점이 있다. CPP는 큐프런시 축의 첫번째 피크(1st rham-

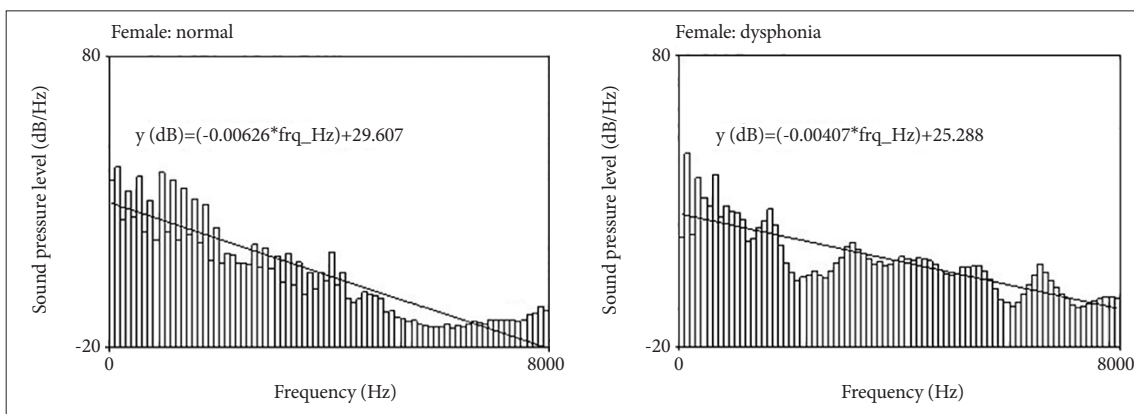


Fig. 10. LTAS object generated with 100 Hz bin width condition (left: normal female, right: female vocal nodule patient). LTAS, long term average spectrum.

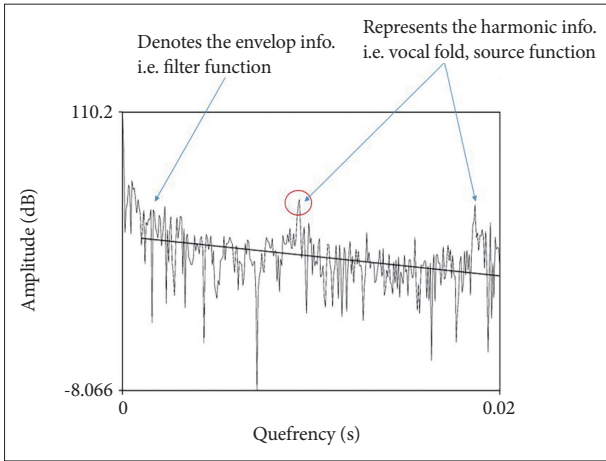


Fig. 11. Cepstral representation of normal voice (cpp: the difference between cepstral peak (red circle) and linear regression line relating quefreny to cepstral magnitude [dB]).

onic; 조화음 배열을 보여주는 상위 큐프런시 영역의 첫번째 피크) 에너지와 큐프런시 축에 분포하는 전체 에너지 선형회귀선과의 에너지 차이값으로 정의된다. Fig. 11에서 붉은 원으로 표시한 부분이 큐프런시 축의 첫번째 피크 즉 첫번째 라모닉(rhmonic)이다. 이 피크값과 선형회귀선 상응값과의 차이 값이 CPP가 된다. CPP는 조화음이 얼마나 구조적으로 잘 조직되어 있는가(harmonic organization)를 알려주는 척도이므로 값이 클수록 주기적인 소리에 가까움을 알 수 있다[23]. 큐프런시 축의 좌측에 성도 공명 성분이 응축되어 있으며 첫번째 피크 이후로 라모닉들이 규칙적으로 배열되어 있다. 이와 함께 큐프런시 축 에너지 값에 연속 윈도우를 취하여 평활(smoothing) 작업을 한 smoothed Cepstral Peak Prominence (CPPs) 변수도 사용 빈도가 높다. Praat에서는 파워스펙트럼 객체와 연동된 동적메뉴 중 smooth..를 활성화하여 구할 수 있다. 대체적으로 스무딩된 결과의 CPP 값이 작게 출력된다.

파워캡스트럼 메뉴와 연동된 query 메뉴에서 구할 수 있는 캡스트럼 변수는 대략 다섯개 정도가 있다. 첫째는 peak 값이다. 기본주파수(1st harmonic)에 해당하는 첫번째 피크 에너지 값(dB)을 말한다. 두번째 quefreny of peak는 그 첫번째 피크값에 대응하는 큐프런시 값을 말한다. 단위는 시간(second)이며 출력된 정보창에 이 시간값에 대응하는 주파수 값도 같이 명시되어 있다. 세번째가 CPP다. 네번째는 trend line slope로 큐프런시 값에 따라 분포하는 에너지 값들을 수학적으로 모델링하여 계산한다. 다섯번째는 이 기울기선이 y 축 즉 에너지축과 만나는 절편(intercept) 값이다. 마지막으로 RNR이 있는데 섭동변수의 HNR과 비슷한 성격의 변수다. 몇 연구에서 머신러닝과 관련하여 유용한 변수임이 밝혀져 있다[19,24]. 이 변수는 가장 최근 Praat 버전 6.2.14에서는 일

괄적으로 10000으로 출력되므로 주의해야 한다. RNR을 다루고자 한다면 이전 버전을 이용해야 한다.

이들 변수 중 가장 보편적이며 많은 연구에서 다루고 있는 CPP의 경우 변수값 출력을 위해서 다음에 제시하는 몇가지 중요한 인수(arguments)들을 필드에 채워야 한다. 1) Search peak in pitch range는 피크값 추적을 위해서 큐프런시 범위(band)를 한정하는 옵션이다. 이 밴드의 하위 큐프런시 값은 $1/\text{pitchCeiling}$ 으로, 상위 큐프런시 값은 $1/\text{pitchFloor}$ 로 규정되므로 성인 남성의 데이터에 주로 사용되는 pitchCeiling 300 Hz라면 하위 큐프런시가 0.0033 sec가 된다. 디폴트는 60-333.3 Hz로 설정되어 있는데 성인 남성용 75-300 Hz, 성인 여성용 100-500 Hz, 어린이 120-600 Hz로 바꿔 설정하더라도 출력되는 CPP 값에 변화는 없다. 따라서 디폴트 세팅을 그대로 사용하면 된다. 2) Interpolation은 피크를 찾는 방식을 5개의 interpolation 옵션으로 결정하게 한다. None, parabolic, cubic, sinc70, sinc700 중에서 cubic과 sinc70은 상대적으로 정밀도가 높다. Cubic은 4개 포인트를 대상으로 수행되며 sinc70이나 700은 70이나 700개 정도의 많은 포인트가 알고리즘에 고려된다. 디폴트 세팅은 세번째에 해당하는 cubic이다. 3) Trend line quefreny range(s)는 큐프런시 값과 상응하는 에너지 값들을 선형회귀선으로 모델링할 때 미리 결정해주는 큐프런시 범위를 말한다. Hillenbrand 등[22]은 매우 낮은 하위 큐프런시 영역의 간섭을 견제하기 위하여 시작 포인트를 0.001초로 결정하였다. 44100 Hz로 샘플링된 음성 파일이라면 큐프런시 범위가 1.486초 정도까지 표현되며 50000 Hz 샘플링된 음성 파일에서는 2.621초 정도까지 분석된다. 디폴트 설정은 0.001-0.05초로 되어 있으며 대개의 피크값이 0.003-0.02초 사이에 존재하므로 디폴트 설정을 준수하면 된다. 4) Trend type은 큐프런시에 따른 에너지 분포를 기울기로 모델링할 때 Hillenbrand 등[22]의 초기 모델처럼 선형직선(straight line)으로 할 것인가 데이터를 좀 더 충실하게 추종하는 지수적 감쇠(exponential decay)형을 선택할 것인가에 대한 옵션이다. 현재 버전(6.2.14 [as of May 24, 2022]) 디폴트는 exponential decay로 설정되어 있다. 5) Fit method는 큐프런시에 따른 에너지 분포를 직선이나 곡선으로 모델링할 때의 계산 방식을 말한다. 디폴트는 Robust slow로 선택되어 있는데 이 방식은 Theil[21]의 Robust 알고리즘을 응용한 것이다. 옵션에는 Theil[21]의 Robust 방식, 전통적 방식인 최소제곱법(least squares), 그리고 Theil[21] 알고리즘의 변형인 Robust slow가 있다. Theil[21]의 알고리즘은 전체 N개의 데이터 포인트에서 N/2개의 포인트쌍을 선택하고 이 갯수만큼 기울기를 계산한다. 최종 기울기 값은 N/2개로 계산된 기울기 값 중 중앙 값(median)으로 결정한다.

관련 연구로 Massachusetts Eye and Ear Infirmary 영어 데이터를 이용하여 295명의 음성장애 관련 진단을 받은 환자 295명(남성 112, 여성 183)의 음성과 50명 정상 화자(남성 20, 여성 30)의 CPPs 절사값을 조사한 결과 모음 연장발화는 14.45 dB, Rainbow 문장에서는 9.33 dB를 보고한 케이스가 있고[25] 한국어의 경우 기능적 음성장애 환자 102명(59.4±12.5세), 정상 집단 59명(52.9±9.1세)의 모음 연장발성 2초를 대상으로 CPPs를 측정하여 기능적 음성장애 집단 11.1 (±2.7) dB, 정상 집단 14.7 (±1.0) dB, 정상과 비정상 집단을 구분하는 절사값 13.84 dB가 보고되어 있다[20]. 음성장애 환자로 성문암 진단을 받은 한국인 남성 피험자를 대상으로 Praat을 이용하여 CPP를 측정한 선행 연구에서는 /아/ 모음 연장발성 2초 정도에서 9.41 (±4.02) dB를 보고한 경우도 있다[26].

성대결절 혹은 성대폴립 진단을 받은 20대 성인 56명(남성 16, 여성 40)을 대상으로 Praat을 이용하여 CPP와 CPPs를 측정한 결과 /아/ 연장발성 3초 정도에서 CPP는 25.27 (±4.21) dB, CPPs는 18.64 (±3.23) dB로 측정되었다(데이터 이용 측정)[19]. CPPs의 이 결과는 위에서 언급한 절사값 13.84 혹은 14.45 dB를 넘어서는 것인데 다른 선행 연구에서 제시한 음성 장애 환자(n=100)의 CPPs 평균값 17.49 (±1.52) dB와는 서로 비교할 만하다[27]. 청지각 테스트 결과 피로한 음성으로 평가된 23명의 정상 여성과 피로도가 느껴지지 않는다고 평가된 정상 여성 27명을 대상으로(전체 평균 나이 28.14세) CPP와 CPPs를 측정한 결과 /아/ 연장발성 2초에 대해서 CPP는 집단간 통계적으로 차이가 없었으며(피로함: 26.55 (±3.92) dB, 피로하지 않음: 27.69 (±3.42) dB, p=0.125), CPPs는 집단간 유의한 차이가 있었다(피로함: 16.82 (±3.26) dB, 피로하지 않음: 18.55 (±2.82) dB, p<0.01, 데이터 이용 측정)[16].

결 론

임상현장 특히 이비인후과와 관련된 분야에서 실용적으로 사용할 만한 Praat 음향변수들을 살펴보았다. Praat이 채택하고 있는 waveform-matching 알고리즘이 1% 이하 소음합성 조건에서도 peak-picking 방식에 비해 안정적임을 알게 되었다. 음향음성학, 음성공학, 음성의학 등 음향학을 다루는 제반 분야에서 dysphonia severity index (DSI)로 일컬어지는 음성장애 진단용 지수를 개발하려는 노력이 1990년대 후반부터 지금까지 이어져오고 있으며, 이 노력의 과정에 Praat의 업데이트가 주요한 기여를 해왔다. 이 분야의 개척자적인 Maryn의 저작들[28,29]을 보면 Maximum Phonation Time, FO_high, Intensity_low, Jitter 등만 포함시킨 초기 DSI에서 2008년에 와서는 CPPs와 spectral tilt 등을 포함하여 Acous-

tic Voice Quality Index (AVQI)로 업그레이드시킨다. 그러나 2008년의 AVQI는 공식(formula)에 포함된 CPPs 계수가 Praat에서 측정한 값이 아니고 speech tool[30]을 이용했으므로 순수하게 Praat 내 변수로 완성된 공식이 아니다. Spectral tilt의 개념도 최근 업데이트를 반영하지 못한 것이므로 수정이 필요하다. 연구자들의 창의적 아이디어와 사용된 데이터의 특성, 데이터 사이즈와 연동된 새로운 DSI용 공식이 앞으로 많이 나올 것이다. Praat의 우수성과 관련하여 CPPs의 성능을 KayPENTAX사 CSL의 Analysis of Dysphonia in Speech and Voice (ADSV) 프로그램의 CPPs와 비교한 Sauder 등[27]에서는 음성장애 진단 정확성의 관점에서 Praat CPPs는 82%의 분류정확도를 보인 반면 ADSV는 75%를 보였고 ROC 곡선의 절사값 추정 정확도를 의미하는 area under the curve는 Praat CPPs가 0.91, ADSV는 0.81로 Praat의 알고리즘이 더 우수함을 입증하고 있다.

중심 단어: 프라트; 섭동; 스펙트럼; 켈스트럼.

Acknowledgments

None

Conflicts of Interest

The author has no financial conflicts of interest.

REFERENCES

- Boersma P, Weenink D. Praat: doing phonetics by computer (version 6.1.16 & 6.2.14) [Computer program]. Amsterdam, the Netherlands: University of Amsterdam, Phonetic Sciences Department [updated 2022 May 25; cited 2022 July 4]. Available from: <http://www.praat.org/>.
- Borden G, Harris K. Speech science primer: physiology, acoustics, and perception of speech. Baltimore, MD: Williams & Wilkins;1984.
- Escudero P, Boersma P, Rauber AS, Bion RA. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. J Acoust Soc Am 2009;126(3):1379-93.
- Yoon TJ, Kang YJ. Monophthong analysis on a large-scale speech corpus of read-style Korean. Phonetics Speech Sci 2014;6(3):139-45.
- Kim JY, Seong CJ. The change of vowel characteristics for the Dysarthric speech along with speaking style. Phonetics Speech Sci 2016; 8(3):51-9.
- Yang B. Development of vowel normalization procedures: English and Korean [PhD Dissertation]. Austin (TX): University of Texas at Austin;1990.
- Fry DB. The physics of speech. Cambridge: Cambridge University Press;1982. p.71-81.
- Park JY, Seong CJ. The implementation of Korean adult's optimal formant setting by Praat scripting. Phonetics Speech Sci 2019;11(4): 97-108.
- Park JY, Seong CJ. The implementation of children's automated formant setting by Praat scripting. Phonetics Speech Sci 2018;10(4):1-10.
- Neel AT. Vowel space characteristics and vowel identification accuracy. J Speech Lang Hear Res 2008;51(3):574-85.
- Sapir S, Ramig LO, Spielman JL, Fox C. Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. J Speech Lang Hear Res 2010;53(1):114-25.

12. Skodda S, Visser W, Schlegel U. Vowel articulation in Parkinson's disease. *J Voice* 2011;25(4):467-72.
13. Boersma P. Should jitter be measured by peak picking or by waveform matching? *Folia Phoniatri Logop* 2009;61(5):305-8.
14. Ko HJ, Woo MR, Choi YL. Comparisons of voice quality parameter values measured with MDVP, Praat, and TF32. *Phonetics Speech Sci* 2020;12(3):73-83.
15. Seo YJ, Shin JY. Acoustic characteristics of the sustained vowel phonation according to age groups. *Phonetics Speech Sci* 2018;10(4):67-76.
16. Jeon HY. A study on the voice fatigue and acoustic characteristics of speech-language pathologists [dissertation]. Daejeon: Chungnam National Univ.;2020.
17. Forrest K, Weismer G, Hodge M, Dinnsen DA, Elbert M. Statistical analysis of word-initial /k/ and /t/ produced by normal and phonologically disordered children. *Clin Linguist Phon* 1990;4(4):327-40.
18. Awan SN, Roy N. Toward the development of an objective index of dysphonia severity: A four-factor acoustic model. *Clin Linguist Phon* 2006;20(1):35-49.
19. Lee IA, Seong CJ. Acoustic characteristics and classification variables of patients with breathy voice. *Journal of the Linguistic Society of Korea* 2020;88:115-34.
20. Kim GH, Lee YY, Bae IH, Lee JS, Lee CY, Park HJ, et al. Acoustic analysis and auditory-perceptual assessment for diagnosis of functional dysphonia. *J Clin Otolaryngol Head Neck Surg* 2018;29(2):212-22.
21. Theil H. A rank-invariant method of linear and polynomial regression analysis. In: Koninklijke Nederlandse Akademie van Wetenschappen, editor. Proceedings of the section of sciences. Amsterdam: North-Holland;1950. p.1397-412.
22. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech Hear Res* 1994;37(4):769-78.
23. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *J Speech Hear Res* 1996;39(2):311-21.
24. Kim NS, Seong CJ. The acoustic characteristics and classification variables of two hyponasal groups. *Journal of the Linguistic Society of Korea* 2017;78:31-61.
25. Murton O, Hillman R, Mehta D. Cepstral peak prominence values for clinical voice evaluation. *Am J Speech Lang Pathol* 2020;29(3):1596-607.
26. Kim GH, Lee YW, Park HJ, Bae IH, Kwon SB. A study of cepstral peak prominence characteristics in ADSV, SpeechTool and Praat. *J Speech Lang Hear Disord* 2017;26(3):99-111.
27. Sauder C, Bretl M, Eadie T. Predicting voice disorder status from smoothed measures of cepstral peak prominence using Praat and analysis of dysphonia in speech and voice (ADSV). *J Voice* 2017;31(5):557-66.
28. Maryn Y, Corthals P, Van Cauwenberge P, Roy N, De Bodt M. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *J Voice* 2010;24(5):540-55.
29. Maryn Y, Morsomme D, De Bodt M. Measuring the dysphonia severity index (DSI) in the program Praat. *J Voice* 2017;31(5):644.e29-40.
30. Hillenbrand J. SpeechTool (Ztool), Version 1.56 [Computer program]. Kalamazoo, MI, USA [updated 2006; cited 2022 July 4]. Available from: <http://homepages.wmich.edu/~hillenbr/>.