

Research on Construction of the Korean Speech Corpus in Patient with Velopharyngeal Insufficiency

Ji-Eun Lee¹, Wook-Eun Kim², Kwang Hyun Kim³, Myung-Whun Sung³ and Tack-Kyun Kwon³

¹Department of Otorhinolaryngology, Chosun University College of Medicine, Gwangju; and

²Departments of Biomedical Engineering, ³Otorhinolaryngology, Seoul National University College of Medicine, Seoul, Korea

구개인두부전증 환자의 한국어 음성 코퍼스 구축 방안 연구

이지은¹ · 김욱은² · 김광현³ · 성명훈³ · 권택균³

조선대학교 의과대학 이비인후과학교실¹, 서울대학교 의과대학 의공학과교실², 이비인후과학교실³

Background and Objectives We aimed to develop a Korean version of the velopharyngeal insufficiency (VPI) speech corpus system.

Subjects and Method After developing a 3-channel simultaneous speech recording device capable of recording nasal/oral and normal compound speech separately, voice data were collected from VPI patients aged more than 10 years with/without the history of operation or prior speech therapy. This was compared to a control group for which VPI was simulated by using a french-3 nelaton tube inserted via both nostril through nasopharynx and pulling the soft palate anteriorly in varying degrees. The study consisted of three transcribers: a speech therapist transcribed the voice file into text, a second transcriber graded speech intelligibility and severity and the third tagged the types and onset times of misarticulation. The database were composed of three main tables regarding (1) speaker's demographics, (2) condition of the recording system and (3) transcripts. All of these were interfaced with the Praat voice analysis program, which enables the user to extract exact transcribed phrases for analysis.

Results In the simulated VPI group, the higher the severity of VPI, the higher the nasalance score was obtained. In addition, we could verify the vocal energy that characterizes hypernasality and compensation in nasal/oral and compound sounds spoken by VPI patients as opposed to that characterizes the normal control group.

Conclusion With the Korean version of VPI speech corpus system, patients' common difficulties and speech tendencies in articulation can be objectively evaluated. Comparing these data with those of the normal voice, mispronunciation and dysarticulation of patients with VPI can be corrected.

Korean J Otorhinolaryngol-Head Neck Surg 2012;55:498-507

Key Words Cleft palate · Database management systems · Speech · Velopharyngeal insufficiency.

Received June 15, 2012

Revised July 25, 2012

Accepted July 26, 2012

Address for correspondence

Tack-Kyun Kwon, MD, PhD

Department of Otorhinolaryngology,

Seoul National University

College of Medicine,

101 Daehak-ro, Jongno-gu,

Seoul 110-744, Korea

Tel +82-2-2072-0738

Fax +82-2-745-2387

E-mail kwontk@snu.ac.kr

서 론

구순구개열(cleft palate)은 선천적으로 구개 혹은 입천장이 갈라져서 구강과 비강이 연결된 상태를 말하며, 700명의 신생아 중 1명에서 발생하는 유병률이 높은 질환이다. 정상인이 발성을 할 때 연구개가 비강과 구강을 차단시켜 비음을 막는 것에 반해 구순구개열 환자는 경구개 혹은 연구개가 갈라져 있어

나 연구개가 선천적으로 짧아 구개인두부전(velopharyngeal insufficiency)이 발생하며 성도로부터 나온 공기흐름이 비강과 구강 동시에 공명하게 됨으로써 발성 및 조음장애가 발생한다. 이의 해결 방법으로 언어치료, 구강내보철물 착용, 수술적 치료 등을 시행하나 반복되는 수술과 장기간의 언어치료를 시행함에도 불구하고 실제 임상에서 구개인두부전 환아의 언어장애를 완치시키기는 어려운 경우가 많다. 더욱이 전화와 같

이 서로 얼굴을 마주하지 않고 시행되는 대화에서는 발음 인식률이 현저히 낮아지기 때문에 문제가 심각하다.

정상인의 음성 코퍼스의 연구와 구축 작업은 국내외에서 이미 상당한 수준으로 이루어진 데 반해, 장애인에 대한 음성인식 및 복원 기술에 대한 연구는 개별적 소규모 자료와 경험을 바탕으로 이루어진 것이 대부분이다. 이에 구개인두부전 환아에서 음성·음운 연구의 객관성과 음성치료의 효율성을 높이기 위한 기초적 바탕 자료로서 구개인두부전 환아의 한국어 발화 음성 코퍼스 구축의 필요성이 절실히 요구된다. 코퍼스는 언어학적 연구를 위해 말 또는 글을 모아 놓은 것으로, 특정 발화자에 의해 발음된 특정 언어 음성을 컴퓨터에서 처리할 수 있는 형태로 전자화 한 것이다.¹⁾ 본문에 거론되는 구개인두부전 환아의 한국어 음성 코퍼스라 함은 구개인두부전을 가지고 있는 환아의 한국어 발화를 대용량으로 녹음한 디지털 음성 자료로서, 컴퓨터로 재구현 될 수 있고, 주석과 그 밖의 자료를 동반하여 체계적으로 정리된 음성 데이터베이스를 의미한다.

본 연구는 구개인두부전 환아의 음성 복원 및 인식 기술을 개발하기 위한 가장 기초단계 과정으로서 기본적이고 필수적인 음성데이터를 선정, 수집하여 국내 관련 연구 기관이 공동으로 이용 가능한 음성 데이터베이스(data base, DB) 구축을 목표로 가장 이상적인 녹음 조건을 구성하고자 하였다. 이는 음성학적인 표준 DB보다 구개인두부전 환아에 특화된 음성 DB 및 전사(transcription)시스템을 구축하여 자료관리 및 검색을 손쉽게 하려는 목적이 있다.

대상 및 방법

발화 방식 및 목록 설계

음성자료수집의 전 단계로서, 정상인, 모의환아 및 구개인두부전 환아가 발화할 녹음 목록을 작성하기 위해 발음 오류 유형을 자음과 모음, 음운 변화의 발음을 고려하여 살펴보았다. 구개인두부전이 있는 환아에서 이미 알려진 발화 오류의 유형은 공명장애: 과대비성(hypernasality), 조음오류, 보상장애: 성문파열음, 인두 마찰음, 음성장애: 기식음(단어 앞에 /h/ 삽입되는 기식음), 긴장음, 소음, 쉰 목소리 등이 있다.²⁾ 이들을 발화목록에 포함시켜 발음 교정을 위한 객관적인 자료를 확보하고, 향후 코퍼스 구축을 통한 후속 연구로 세부적인 발음 오류 분석을 실시하여 과학적으로 객관화하고자 하였다.

자연스럽게 발성된 음성데이터의 수집을 위해 발성 리스트를 만들어 참가자들에게 문장을 읽게 하였다. 음성파일의 경우, 다양한 데이터를 얻기 위해 낭독체 및 낭독체 음성에서 나타나지 않는 고유한 표현들을 관찰하기 위해 자유발화형식에

Table 1. Speech materials used in our experiment

모음	아, 이, 우, 야, 예, 위
자음	가, 바, 사, 자, 라, 하, 까, 따, 빠, 싸, 짜, 카, 타, 파
비강음	마마, 나나, 앙앙
낱말	비강음 (-) 바지, 포도, 사과, 빨대, 식탁, 코끼리 비강음 (+) 김밥, 전철, 식빵, 땅콩, 자동차, 장난감
문장	비강음/구강압 자음이 없는 문장 이리로 올라오래요. 파열음으로 구성된 문장 보배 아파 업고 가요. 비강음으로 구성된 문장 맘마 많이 먹는 매미 파찰음으로 구성된 문장 저 차 의사에서 자요. 마찰음으로 구성된 문장 이 새우 짜서 샀어요.
문단	토끼와 거북이 토끼와 거북이가 달리기를 했어요. (비음비율 0%) 토끼는 앞에서 빨리 뛰어 갔고 거북이는 뒤에서 기어 갔어요. 아기(비음비율 11.7%) 아기가 엄마 품에 잠들어 있네요 우리 아기 예쁜 아기 새근새근 잠자요. 엄마(비음비율 34.4%) 엄마는 항상 레몬쨈을 만들어 이모랑 누나랑 나누어 줍니다. 우리 엄마 좋은 엄마

대해 모두 데이터를 수집했다.³⁾ 낭독체 발성 목록은 기존의 구개인두부전 환아의 음성언어치료 목적으로 사용되는 대본을 구개장애인들에게 주로 나타나는 오류를 중심으로 모음, 자음, 비강음, 낱말, 비강음/구강압 자음(파찰음, 파열음, 마찰음과 같이 발화시 높은 구강 압력이 필요 없는 자음)이 없는 문장, 파열음/비강음/마찰음/파찰음으로 구성된 문장, 그리고 비음 문장에 대한 비음도(nasalance)를 알아보기 위해 비강자음 비율이 각각 0%, 11.7%, 34.7%인 토끼 문장, 아기 문장 그리고 엄마 문장으로 구성되도록 하였다(Table 1). 발성 목록을 발화하게 한 이후에는 언어치료사와 함께 일상생활에 대한 이야기를 나누는 방식으로 자유발화 녹음을 진행하였다.

화자 구성 및 수집 환경

음성데이터 수집 대상

구개인두부전의 장애 심각도가 가능한 다양하게 포함되도록 많은 데이터의 음성 코퍼스를 확보하는 것이 중요하다고 생각되어 이전의 수술 받은 경력이나 음성언어치료를 받은 경력에 관계없이 음성언어치료를 받는 만 10세 이상의 구개인두부전 환아를 대상으로 하였다. 음성녹음 과정에 협조가 잘 되고, 발화 목록에 따른 발성이 적절히 되도록 하기 위해 10세라는 나이제한을 두었다. 구개인두부전 환아의 발음에 영향을 미칠 수 있는 나이, 성별, 음성언어치료 기간, 수술시행시기 등의 자료는 저장해두어 향후 분석에 사용하도록 하였으며, 녹음대상자의 선정 및 제외 기준이 되도록 하지 않았다.

모집과정에서 정상 모의 환아와 구개인두부전 환아에게 녹

음의지를 확인한 후 피험자 동의서를 받았다(Informed consent approved by Institutional Review of Board of Seoul National University Hospital, IRB Number: 1103-040-354). 대부분의 발화자들은 구개 구순열 수술 후 언어치료를 위해 외래에 정기적으로 내원하는 환아로 구성되었다.

음성녹음은 크게 정상 모의 환아, 구개인두부전 환아 군으로 나누었다. 정상 모의 환아 군을 포함시킨 것은 같은 발화 내용에 대해서 환아와의 발음 비교 분석을 수월하게 하고 향후 발음 교육에 활용하기 위해서이다. 한국어 표준말을 사용하고 청력장애가 없는 건강한 성인 5명을 정상 모의 환아 군으로 하였고, 이들의 연령은 27세에서 33세까지 평균 28세였다. 구개인두부전 환아는 구개 파열로 이전에 수술을 시행받았거나 혹은 받지 않았더라도 외래추적 관찰하며 음성언어치료를 시행하고 있는 환아를 대상으로 하였고, 이전에 수술을 시행하였지만 발음교정이 불완전하여 재수술을 대기하고 있는 경우나 수술이 예정되어 있는 구개인두부전 환아도 음성녹음을 하였다. 구개 파열 이외의 정신 신체 장애나 발성 및 조음에 문제가 있는 경우는 대상에서 제외하였다.

녹음 기기-3 채널 동시 녹음 장치

실제 임상에서 사용되는 Model 6450 nasometer(Kay PENTAX, Lincoln Park, NJ, USA)는 비음의 정도를 측정하는 검사도구이다(Fig. 1A). 비음-구음 분리판 양면에 설치된 마이크로폰으로 음성신호가 입력되어 시스템 내로 입력이 되면 기기에 의해 측정된 에너지에 의해 비음도를 계산한다. 비음도는 구강 음향학적 에너지와 비강 음향학적 에너지 합에 비강 음향학적 에너지의 비율을 퍼센트로 아래 공식에 의해 계산할 수 있다.⁴⁾

$$\text{비음도(nasalance)} = [N/(N+O) \times 100]$$

N: 비강 음향학적 에너지, O: 구강 음향학적 에너지

본 연구에서는 의공학과에서 Model 6450 nasometer를 모델 삼아 자체 제작한 비강 측정기를 이용하였는데(Fig. 1B), 이 측정기는 녹음이 가능한 3 채널을 가지고 있다. 구음-비음 분리음 녹음 장치의 구성은 Fig. 1B와 같이 마이크를 두부에 고정

할 수 있게 해주는 headset, 음향 분리판의 위와 아래에 비음과 구음을 따로 수집할 수 있는 2개의 condenser 마이크(Fig. 1C)가 있고, 마이크와 입까지의 거리는 대략 10 cm 정도를 띠었다. 마이크로부터 전달된 음성신호를 컴퓨터 사운드카드의 스테레오 입력으로 받을 수 있을 만큼 큰 신호로 다시 증폭(10 배)하는 증폭기 및 측정 노이즈를 최소화하기 위한 배터리로 구성되어 있다. 이렇게 자체 제작을 한 이유는 구개인두부전 음성의 특성연구를 위한 자료로서 비음과 구음이 분리되어 녹음된 음성(분리음)과 분리되지 않은 음성(비분리음)을 각각의 채널을 통해 동시에 얻는 것이 구개인두부전 음성의 음향학적 특성을 더 자세히 볼 수 있을 것이라는 생각에서이다. 한편 이들 3 채널의 오디오는 단일 녹음 시스템을 이용해 동시에 녹음이 되도록 하였는데, 이는 초기 분리음과 비분리음을 따로 녹음하였을 때 각각 녹음된 데이터의 시간 축의 값이 일치하지 않아 데이터 처리 및 검색 등에서 어려움이 발생하여 데이터베이스 구축 작업에 어려움을 가져온다고 판단되었기 때문이다. 음성은 마이크로폰에 의해 전기신호로 바뀐 후에 PC에서 비강음과 구강음을 각각 그리고 합쳐서 표현된 결과를 얻었다. 비분리음은 통상 녹음 작업에서와 마찬가지로 입으로부터 40 cm 떨어진 위치에 고감도 마이크를 설치하여 측정하였다(Fig. 1D). 마이크로부터 나오는 음성신호를 다목적 USB 녹음장치인 U46XL(SuESI Audiotechnik GmbH, Leonberg, Germany)을 이용하여 주파수 44.1 kHz, 양자화 비트수 16 bit로 디지털화하고 Cubase LE5 소프트웨어를 이용하여 녹음 파일을 얻었다. 높은 주파수로 샘플링을 한 이유는 높게 하더라도 낮은 샘플 주파수로 변환이 가능하고, 가능한 많은 모집단 데이터(raw data)를 구축하려고 하였기 때문이다. 그리고 이 모든 녹음 장비 및 장소는 음성 코퍼스 부가 정보란에 포함되도록 하였다.

녹음 방법

녹음은 주변 환경 소음을 최대한 피하기 위해 외래 진료 후 언어치료실 한 곳을 지정하여 시행하였다. 녹음과정에는 음성 언어치료사, 의공학과 연구원, 이비인후과 의사가 참여하였고,

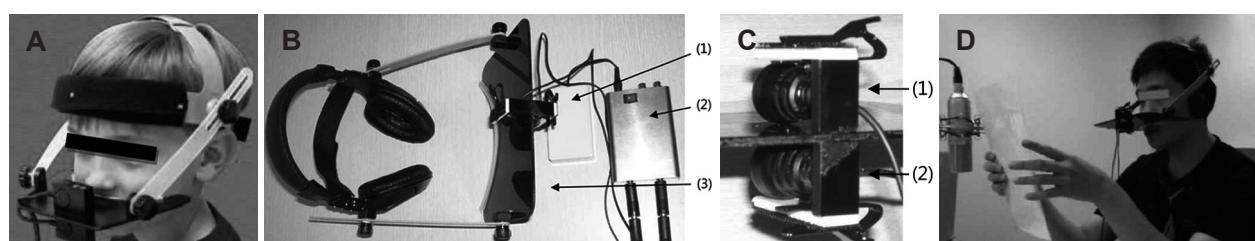


Fig. 1. Recording device. Nasometer II, Model 6450, Kay PENTAX, Montvale, NJ, USA (A). Recording device made by our team. (1) battery, (2) amplifier, (3) laminated plate for separating the nasal/oral sound (B). (1) Microphone for recording the nasal sound, (2) microphone for recording the oral sound (C). Applying of the recording device in patient (D). The device is composed with the headset, laminated plate for separating the nasal/oral sound, 2 condenser microphones and amplifier.

언어치료사의 주도하에 발화자가 긴장하지 않도록 최대한 편한 분위기를 조성하고, 오류 발화의 경우에도 수정을 지시하지는 않았다. 재발화 녹음은 녹음시 실수로 빠뜨린 것만 재녹음을 실시하였다.

정상 모의 환아의 경우 정상상태에서 한 번, 그리고 실험적으로 구개열 발음을 시뮬레이션 하기 위해 같은 별화 목록으로 두 번 더 녹음하였다. 구개열 시뮬레이션에는 3 프렌치(French) 구경 크기의 넬라톤 카테터(nelaton catheter)를 그림과 같이 양측 비강을 통해 넣고(Fig. 2) 긴장도가 없는 상태에서 위치(1)을 모스끼토(mosquito)로 표시해 놓고, 통증을 유발하지 않는 선에서 최대의 긴장도가 생성되는 위치(2)를 표시하였다. 넬라톤 카테터가 (2)의 위치에 있을 때를 구개열 모의 환아 중증(severe) 상태로 하였고, (1)과 (2)의 중간위치에 있을 때 구개열 발음이 녹음된 것을 경도(mild) 상태로 정의하였다.

이전에 비디오 녹음 영상이 완료되어 있는 구개인두부전 환아의 경우 44.1 kHz, 16 bit의 PCM Coding되어 있던 음성을 Audio Extraction 프로그램을 이용하여 분리한 후 16 kHz, 16 bit로 PC 파일화하고 슬라이싱(slicing) 툴을 사용하여 어휘는 각 어휘별로, 문장은 문장별로, 이야기 읽기는 이야기 읽기별로, 대화체는 대화 세트별로 파일화하였다. 음성자료의 체계적인 디렉터리 구조를 설정하기 위해 윈도우 액세스(Access) 파일을 이용하여 여기에 모든 발화자의 부가정보를 기입하였다. 발화자인 구개장애 환아의 나이, 성별, 구개장애 타입분류(type), 첫 수술시기, 수술 종류, 수술 횟수, 수술 이후 추적 관찰 시간, 발화 능력, 언어치료 시행기간 등의 부가 정보를 기입하였다.

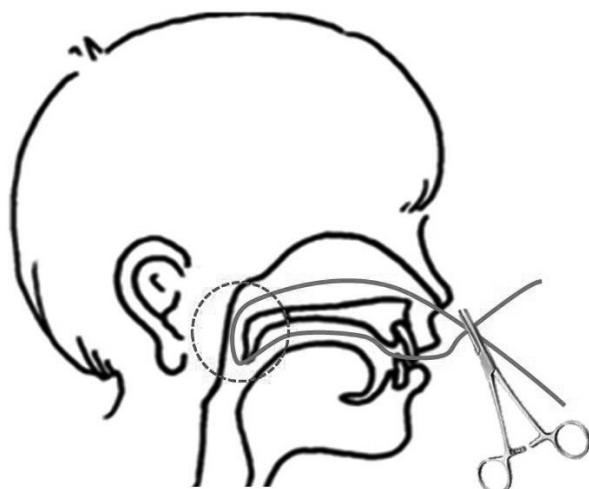


Fig. 2. Illustration of simulated velopharyngeal insufficiency with 3-french nelaton tube. In control group, to induce the velopharyngeal insufficiency, authors inserted 3-french nelaton tube in each nostril through nasopharynx and pulled the soft palate anteriorly in varying degrees. The dotted circle stands for the site of velopharyngeal insufficiency which has been induced.

전사 과정

완료된 녹음데이터에 대해 언어치료사의 주도하에 2명의 비전문가가 참여하여 발화 내용을 그대로 문자로 전사한 후 오조음(misarticulation)이 발생한 부분에 대해 분석 태그를 입력하였다. 발화자의 전체 음성 특성 및 명료도를 7단계로 표현하고 전통적인 조음장애의 분류와 구개열 환아의 특성을 고려하여 생략(omission), 대치(substitution), 왜곡(distortion), 첨가(addition), 과대비음화(hypernasality), 과소비음화(hyponasality), 보상조음화(compensation) 및 비누출(nasal turbulence)의 8개 오조음 패턴으로 분류하여 기록한 후 전사 및 오조음 분석의 원본으로서 PDF 포맷 파일로 저장하였다. 저장된 전사 및 오조음 분석 파일의 일부분을 Fig. 3에서 볼 수 있다.

데이터 베이스의 구축

본 연구에서 설계한 구개인두부전 음성 코퍼스는 Microsoft 사의 MS Office Access 2007을 사용하여 구현했다. 데이터베이스는 patients, records, transcription의 3 테이블로 구성되어 있으며(Fig. 4A), 이들 각각은 기본 키를 이용해 각각이 일대다(one to many)의 대응관계를 맺도록 했다. 이는 한 환아가 여러 번 녹음에 참여할 수 있으며 녹음할 때마다 해당 환아의 새로운 레코드 데이터가 발생할 것이고, 다시 각 음성파일로부터 전사 및 오조음 데이터가 발생할 것이므로 이를 표현하기 위해 가장 적절한 테이블 관계라고 볼 수 있다.

환아 정보를 저장하는 테이블(patients table)(Fig. 4B)에는 이름, 데이터 저장소 이름, 성별, 방언지역, 생년월일 등의 기본 정보를 저장하도록 하였고, 환아 정보 보호를 위해 환아 이름 필드는 최종 데이터베이스에서는 제외되도록 하기로 하였으며, 데이터베이스 제작상의 편의를 위해 임시로 존재하는 필드를 마련하였다. 다음으로 한 환아로부터 얻어진 음성데이터들을

과대비음: moderate to severe	과비
Hoarseness	보상조음(ㄱ, ㄷ, ㅂ, ㅅ, ㅈ)
보상조음 매우 많음	ㅋ, ㅌ, ㅍ, ㅊ, ㅆ, ㅉ
명료도: Intelligible with careful listening	ㅋ, ㅌ, ㅍ, ㅊ)
아, 이, 우, 야, 위, 약, 어, 어, 오, 요, 우, 위	마마, 나나, 양양
가, 다, 바, 사, 자, 라, 하,	바지, 포도, 사과, 빨대, 식탁, 코끼리
까, 따, 빠, 싸, 짜,	김밥, 전철, 식빵, 땅콩, 징둥차, 징난감
카, 타, 파, 차	보(ㅂ, ㅈ, ㅍ, ㄷ, ㅅ, ㄱ, ㅋ, ㅌ, ㅊ, ㅍ, ㅌ, ㅊ, ㅍ, ㅊ)

Fig. 3. Analysis pdf file made by speech therapist, which is composed of transcription script and misarticulation examples. General character and clearance of the voice are scored on a 7-point scale. Misarticulation is categorized and recorded as 8 subgroups-omission, substitution, distortion, addition, hypernasality, hyponasality, compensation and nasal turbulence.

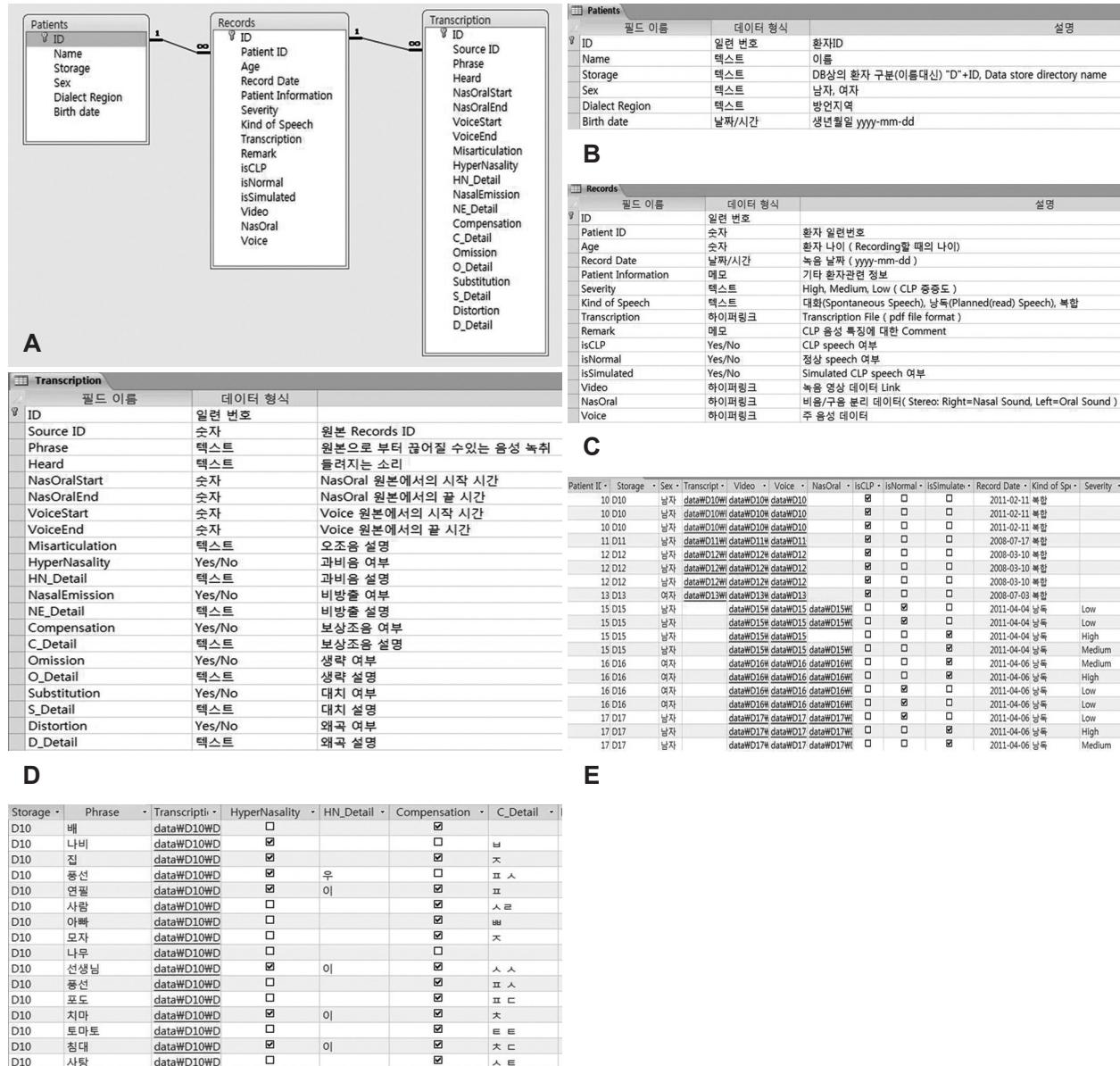


Fig. 4. Database table (MS office Access 2007). Relation of the database table (A). Design of the patients table (B). Design of the records table (C). Design of the transcription table (D). Illustration of voice data entering (E). Storage of database in transcription table (F). The database table is consisted with three components-pateints, records and transcription tables, each table had several-for-one relationship. The recorded voice files are stored in directories of their own, and are hyperlinked to the lists on the record table.

기록하는 테이블(record table)(Fig. 4C)은 녹음 당시의 나이, 날짜, 환아의 병적 정보, 중증도, 발화방법(대화/남독/복합), 음성파일의 전사기록, 음성의 특성, 녹화 영상 파일 유무 등을 기록하는 필드를 만들어 놓았다. 구강/비강음 분리 및 비분리 음성데이터 파일에 대해 하이퍼링크 시켜놓아 즉석에서 파일을 열어볼 수 있도록 만들었다. 계속하여 transcription 테이블(Fig. 4D)에는 녹음된 음성데이터들을 분석하여 각각의 특징들을 기술하고 해당 발음이 음성데이터의 어느 부분에 위치하는지를 알려주는 시간 인덱스를 저장한다. 언어치료사에 의

해 해석된 전사 내용을 데이터베이스에 기록하는 Table을 Fig. 4D에 나타내었고, 전사에서는 오조음으로 과비음, 보상조음, 비방출, 생략, 대치, 왜곡 등의 존재 유무와 해당 음성 파일의 어느 시간에 위치하는지를 가리키는 Time V(비분리음 시간 위치), Time N(분리음 시간 위치), 구음/비음 분리 음성데이터 존재 여부, 오조음 존재 여부 및 오조음에 대한 설명을 입력할 수 있게 하였다(Fig. 4E and F). 녹음된 음성데이터들은 wav 파일로 환아 고유의 디렉토리에 저장되며 records 테이블에는 해당 파일을 하이퍼링크 시켜 놓아서 자료의 추출을 용이하도록

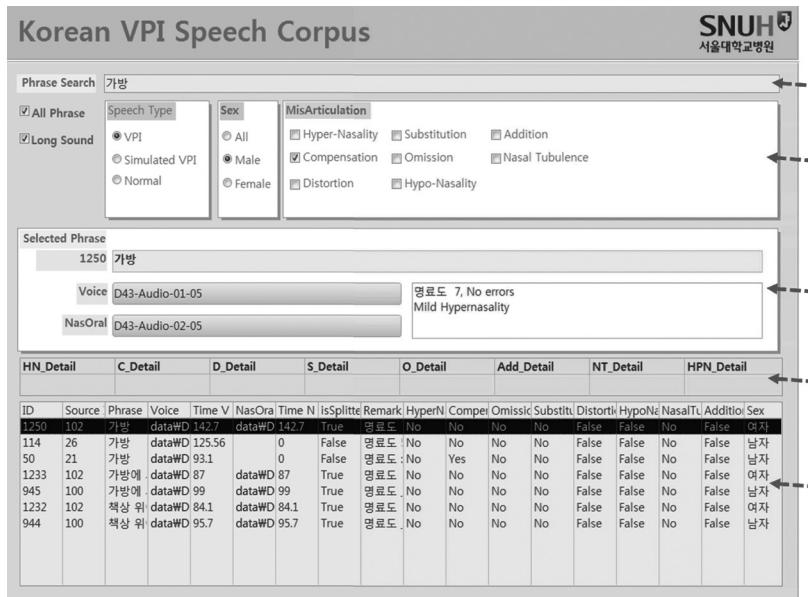


Fig. 5. Query form of corpus for searching phrase and interface between Microsoft Access 2009 and Praat program. (1) entering a search phrase, (2) conditional search, (3) information about the selected phrase, (4) information about the misarticulation of the selected phrase (5) search results. There are five panels in query form of corpus which facilitate the searching process.

하였다.

쿼리 폼(Query form) 작성

Fig. 5과 같이 검색어구, 발화 형태, 성별, 오조음 조건 등을 입력하여 데이터베이스 상에서 원하는 검색어구가 포함된 모든 데이터를 검색 결과 창에 표시되도록 하는 쿼리 폼을 작성하였다. Fig. 5의 ①, ②에 원하는 검색 조건을 선택하면 ⑤에 검색어구를 포함하고 있는 모든 전사 입력된 결과가 나타난다. 이 중에 한 문구를 클릭하면 ③열에 Voice 혹은 NasOral 음성파일 데이터가 위치한 디렉토리 및 파일 이름 확인이 가능하다. 문구가 적힌 네모칸은 각각 클릭하면 데이터베이스의 해당파일로 링크되어 Praat script를 호출하는 임시 명령이 생성되고, Praat 프로그램이 실행되어 저장되어 있던 검색어구 및 오조음 분석 내용 결과를 화면에 표시해준다.

장애 음성 검색 및 데이터 분석 소프트웨어 인터페이스

Access 프로그램을 이용하여 데이터 파일링을 한 이후 음성 데이터의 검출과 추출을 용이하게 하기 위해 음성분석 전문 프로그램을 Praat를 인터페이스 하였다. Praat는 프로그램을 개발하기 전에 어떤 과정과 어떤 음성 파라미터를 이용하는 것이 적절한지 그래프 등을 이용해 결과를 미리 확인해 볼 수 있는 도구로, 피치분석, 스펙트로그램 분석, 포먼트(formant) 분석, 강도 값 측정 및 변환 등 음성신호 처리를 위한 대부분의 연구를 진행할 수 있다.⁵⁾ 특히 사용자가 직접 분석 파라미터를 수치로 입력할 수 있는 대화창을 통해 음성데이터 파형과 오조음의 형태 및 위치를 전문적으로 처리할 수 있는데, transcription 테이블에 입력되어 있는 모든 데이터들을 Praat script를 만들고 데

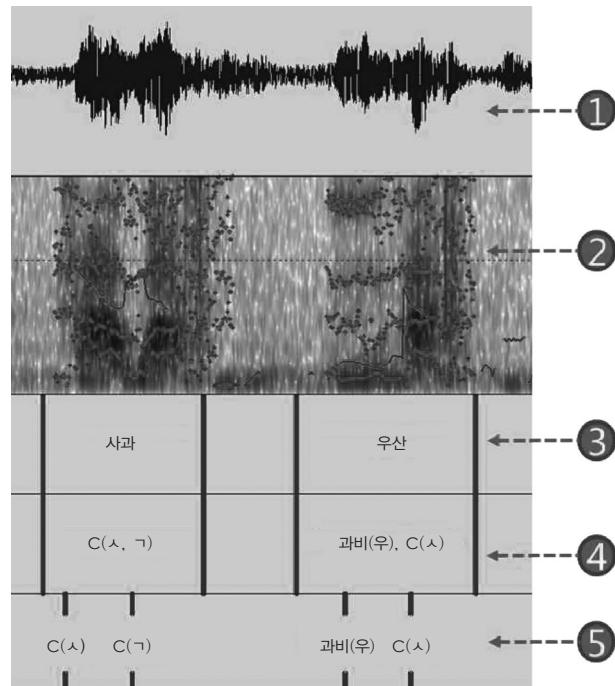


Fig. 6. Text Grid file showing misarticulation and time index. (1) wave patterns of voice data, (2) spectrogram, (3) spoken letters, (4) analysis of misarticulation, (5) time index of misarticulation. Each spoken letter was analyzed and transcribed on the Praat text grid file.

이터페이스 내부에 이와 관련된 프로그램 Visual Basic for Application 코드를 이식하였다. 이렇게 함으로써 데이터 베이스 검색으로부터 전달되는 검색어구와 시간 인덱스가 각각의 음성데이터 및 전사 분석내용을 코딩하여 Praat의 Text Grid 파일을 호출할 수 있게 된다. Text Grid 파일에는 전사 및 오조음 분석 태그를 달아서 ③ 발화 음성(사과, 우산), ④ 오조음 분석

(C: compensation, 과비; 과비음) 및 ⑤ 오조음이 발생한 시간 위치를 표현하도록 하였다(Fig. 6).

결 과

음성녹음 결과

모의 발화자 7명(남 5, 여 2)과 구개인두부전 발화자 26명(남 18, 여 8)에서 전체 94개의 발화 음성파일을 녹음하였다. 모의 발화자는 표준어를 구사하는 25~31세의 성인남녀로 구성되었고, 구개인두부전 발화자의 경우 구순구개열 환아 15명, 구개열 환아 5명, 구개수열 환아 1명, 구개인두부전만 있는 경우가 3명이었다.

기존의 언어치료사가 가지고 있던 발화 데이터를 제외하고 나머지는 모두 3 채널로 비음/구음의 분리음 및 비분리음의 음

성데이터를 획득하였다.

Praat를 이용한 구개인두부전 환아의 음성데이터 모의분석

Fig. 7에 수술 후 구개인두부전이 남아 음성치료를 시행하고 있는 20세 남자 환아의 “이리로 올라오래요”의 음성 스펙트로그램을 나타내었다. 위에서부터 각각 구강음/비강음/비분리음의 파형을 나타낸다. “이리로 올라오래요” 문장의 “이리” 부분에서 비음에너지(B)가 구강 에너지보다(A) 크게 나타나는 것을 관찰할 수 있는데, 이는 구개인두부전 음성의 모음/이/에서 비음이 크게 나타난다는 사실과 부합된다. 그림의 왼쪽 세로축은 위: 음성신호 크기 및 아래: 포먼트(formant) 주파수이고, 오른쪽 세로축은 피치주파수 및 각 아래칸 내의 연결선은 음의 피치(pitch)를 나타낸다.

모의 환아의 음성데이터 분석

모의 환아의 음성데이터 중 비음 17.4% (“토끼와 거북이가 달리기를 했어요”)와 54.0%의 문장 (“엄마는 항상 레몬쨈을 만들어”)에 대하여 비음과 구음의 각 구간 에너지를 Praat를 이용하여 얻은 후 비음도를 계산해 보았다. 문장의 종류에 따라 비음값이 예상 %까지 미치지는 못하였으나, 정상 상태인 경우와

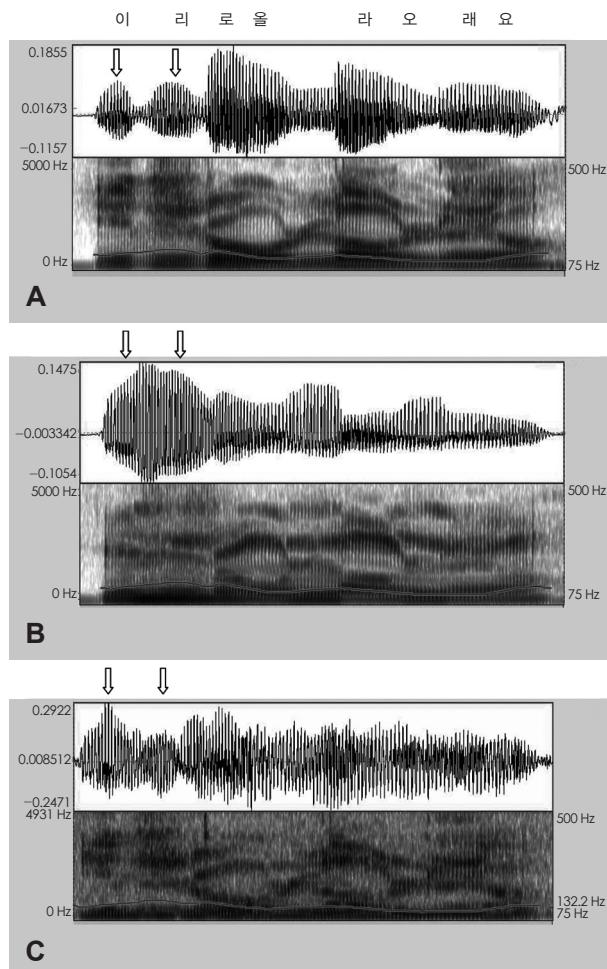


Fig. 7. Comparison of the wave form and spectrogram between the oral/nasal sound and the non-separating sound by using the sentence “이리로 올라오래요”. Oral sound (A). Nasal sound (B). Non-separating sound (C). The wave form showed the highest nasal energy in “이리” phrase (arrows), while pronouncing the sentence “이리로 올라오래요”.

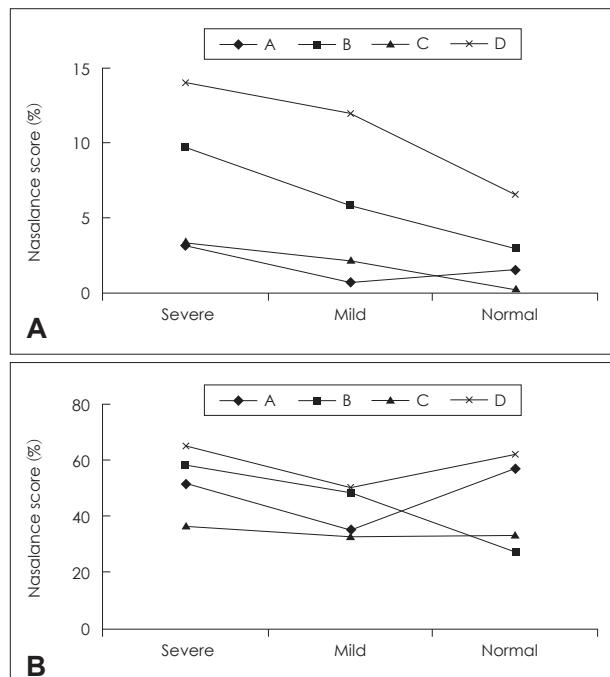


Fig. 8. Nasalance of the simulated velopharyngeal insufficiency speech. Nasalance measured in the simulated patient by using the sentence containing a percentage of 17.4% nasalance (A). Nasalance measured in the simulated patient by using the sentence containing a percentage of 54.0% nasalance (B). The more severe the velopharyngeal insufficiency, the nasalance calculated by the Praat program was higher (A, B, C and D in figure denotes different simulated persons).

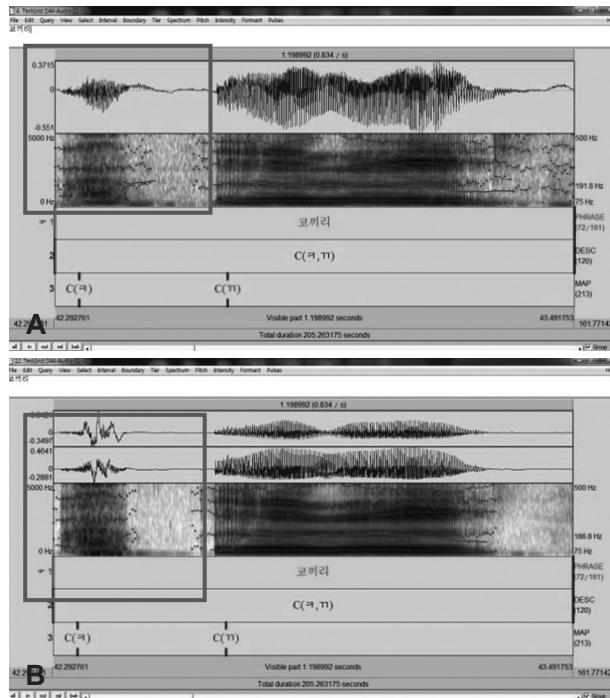


Fig. 9. Wav and Text Grid file showing the spectrogram of the phrase “코끼리” (patient with velopharyngeal insufficiency). Non-separating voice data (A). Voice data recorded separately via two channel; naso and oral (B). Non-separating and naso/oral separating voice data and misarticulation analysis file which was interlocked with Praat program in velopharyngeal insufficiency patient pronouncing the “코끼리” were demonstrated (A). There was compensation in consonants such as “ㅋ” and “ㅌ” (B). In red box, oral energy in upper panel and nasal energy in lower panel were demonstrated separately.

비교해 십각도에 따라 비음도가 커지는 것을 확인할 수 있었다(Fig. 8).

모의 환아와 구개인두부전 환아의 발화음성 비교

본 프로그램을 이용하면 정상 모의 환아와 구개인두부전 환아에서 발화되는 어구를 동시에 검색 및 출력하여 발화음성 파형 및 스펙트로그램의 비교가 가능하다. Fig. 9는 실제 “코끼리”라는 어구에서 나타난 보상조음 및 과비음 형태의 오조음을 가지는 구개인두부전 남자 환아를 검색 후 Praat 상으로 출력하게 하여 해당 어구의 데이터 파형과 Text Grid 파일이 함께 표시되게 한 것이다. “ㅋ”과 “ㅌ”에서 과대보상이 나타나고 있고, 과대보상이 나타나는 시간을 위치상으로 표현하였다. (B)에서는 박스에 보이는 바와 같이 잇줄의 구음과 아랫줄의 비음에너지가 나눠 표현되는 것을 볼 수 있다.

Fig. 10은 같은 발음을 정상인에서 했을 때 비분리음(위)/분리음(아래) 파형을 나타낸 것이다. 그림내의 빨간 사각형은 “코-”가 발화될 때의 비분리음 및 분리음의 음성파형 상 전체 에너지와 비강/구강 에너지 파형을 나타낸다. 구개인두부전 환아와 정상인에서 두 개의 파형이 다르게 나타나는 것을 쉽게 관

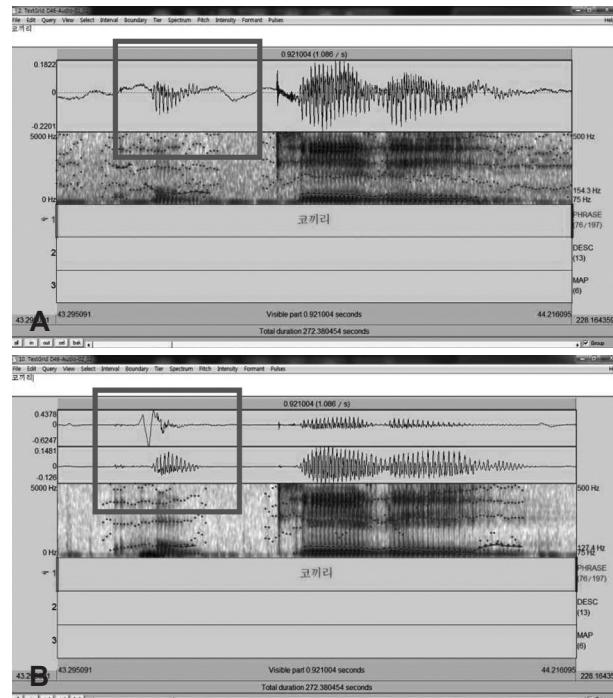


Fig. 10. Wav and Text Grid file showing the spectrogram of the phrase “코끼리” (control). Non-separating voice data (A). Voice data recorded separately via two channel; naso and oral (B). Non-separating and naso/oral separating voice data and misarticulation analysis file which was interlocked with Praat program in simulated patient pronouncing the “코끼리” were demonstrated. When compared with the Fig. 9, the wave form was different from that of velopharyngeal insufficiency patient.

찰할 수 있다.

고 찰

음성언어의 정보화

음성언어를 자료화하기 위해서는 녹음과 전사 과정을 거쳐야 하기 때문에 문자언어를 자료화하는 일에 비하여 많은 시간과 노력 그리고 비용이 필요하다. 그럼에도 불구하고 음성언어가 가지는 특성은 음성언어가 정보화되어야 하는 이유를 제공한다. 음성 자료는 발화자에 따른 다양한 음성적 변이형을 가지게 되는 특징이 있다. 또한 발화자의 실수가 발생한 경우, 실수 그대로 자료의 일부가 되고, 실수를 일으킨 형태와 녹음 과정에서 만일 교정이 일어났다면 교정된 형태까지도 좋은 음성 자료가 된다. 그리고 음성자료는 발화된 음성 신호 자체뿐만 아니라 성대 진동 여부를 보여주는 후두조영술(laryngography), 비음측정도, 조음기관의 위치를 관찰하기 위한 자기공명장치, 방사선(X-ray) 자료 등의 다양한 신호가 모두 음성언어 자료가 될 수 있다. 이러한 모든 과정의 최종 결과물은 단순 음성언어의 녹음 자료가 아닌 발화자에 의해 발음된 음성을 컴퓨터가 읽을 수 있는 형태로 디지털화 된 후 자료에 대한 정보가 세세

하게 밝혀져 널리 공유되어 재사용 될 수 있는 음성언어 코퍼스의 구축이다.⁶⁾

2003년 국립국어원에서 제작되어 2005년 개발 발표된 '서울 말 낭독체 발화 말뭉치'는 2대 이상 서울 경기지역에 거주해 온 서울말 화자 120명의 낭독체 말소리를 담아 만든 것으로 소설, 수필, 논설 등 다양한 형태의 글을 편하게 읽은 음성 파일 약 88800여 개가 수록되어 있다. 이것은 CD로 제작되어 일반인에게 널리 무료 배포되어 사용되고 있는데, 정상 발화자의 다량 wav 파일을 담고 있다는 점에서는 유용하나 어구 검색창의 부재로 인해, 원하는 어구를 빠르게 찾아내고 분류하거나, 같은 어구를 다른 조건의 발화자 파일에서 비교 분석하거나 혹은 두 개 이상의 wav 파일을 동시에 비교작업 하기에 불편함이 있다. 본 연구에서 고안한 코퍼스는 이와 같은 단점을 극복하기 위해 Access 데이터베이스에서 검색조건을 설정 가능하게 하여, 어구의 검색 및 음성파일의 입력 및 출력을 빠르게 하였다.

구개인두부전 환아에서 음성 코퍼스 구축이 가지는 의의

특정 조건을 만족하는 발화자의 음성언어 자료 구축은 음성 공학적 분야 혹은 음성언어의 전사 코퍼스를 구축하는 일반적인 코퍼스 구축의 두 가지 명분에 더하여 임상의학적인 면에서의 필요를 충족시킬 수 있다. 산전 진단의 기술이 발달함에 따라 구개구순열과 같은 질병의 초기 진단이 용이해지면서, 구개구순열을 지닌 상태로 성장 후 이비인후과를 찾는 환아보다는 생후 1~2년에 성형외과, 구강외과 등에서 수술을 받은 후 완전히 교정되지 않은 발성 및 조음장애의 교정을 위해 이비인후과를 방문하는 경우가 대부분이다. 따라서 현재 우리가 임상에서 실제 경험하는 환아는 수술 후 혹은 다른 여러 일차 치료가 시행된 경우가 많고, 나이 또한 매우 어려서 명확하고 의미전달이 가능한 발음을 시켜보기 어려운 경우가 많다. 음성 코퍼스 구축에서 모의 인두구개부전 발화를 포함시킨 것은 이러한 연유이며, 정상음과 구개인두부전 음성의 심도 있는 분석을 통해 구개인두부전 음의 발생과정의 음성, 음운학적 연구의 객관성을 높일 수 있다.

한편 불완전 교정 환아에서는 음성언어치료사의 도움으로 어느 정도 교정치료가 가능하지만, 일부에서는 음성치료만으로는 한계가 있어 재수술 혹은 다른 방법의 치료를 모색해야 하는 경우가 있다. 이러한 상황에서 음성 코퍼스를 이용하여 장애 환아 발음의 명료도 혹은 남아있는 장애 정도가 어느 단계인지 정확한 진단이 가능하고, 음성치료로 성취 가능한 음성언어 단계가 어디까지인지 명확히 서술된다면 환아 및 보호자와의 의사소통의 향상을 가져올 수 있다. 뿐만 아니라, 정상인의 음성자료와 함께 비교 분석이 가능하므로 구개인두부전 환아의 조음 치료 성공 및 추적 관찰의 지표로써 도움을 줄 수

있다.

본 연구의 궁극적인 목표는 구개인두부전 환아의 음성언어 코퍼스 구축을 넘어, 음성 신호 처리 및 인간-기계(Man-machine) 인터페이스의 개발을 위해 의공학자, 언어치료사, 임상의가 함께 환아에서 나타나는 공통적이고 대표적인 조음의 어려움과 경향성을 찾아내어 이를 다시 환아의 진단 치료 및 재활에 유용한 데이터로 사용할 수 있도록 하는 것이다. 결국, 음성언어 자료의 구축은 언어학적 공학적 연구뿐만 아니라 음성 합성 기술의 응용을 가능하게 하고, 더 나아가 음성 인식, 음성 합성 및 자동 통역을 이용하여 구개인두부전의 치료에 큰 진척을 가져올 수 있도록 하는 기반이 되고자 하는 데 궁극적인 의의가 있다고 하겠다.

음성언어 코퍼스의 임상 적용성

현재까지 음성언어 장애가 있는 환아를 대상으로 개발되어 실용화 단계에 있는 장치로써 음성의 진폭을 크게 하는 기능에 부가하여 음성을 보다 선명하게 들릴 수 있도록 하는 어음향상 시스템(speech enhancer) 기술을 포함한 장치가 있다. 음성을 합성하여 듣는 사람이 정상인과 비슷하게 음성을 이해할 수 있도록 개발된 시스템으로 전화기와 연결하여 사용할 수 있도록 개발된 인터페이스이다. 뇌졸중(stroke), 뇌성마비(cerebral palsy), 음성장애(voice disorders), 다발성 경화증(multiple sclerosis), 척추 손상(spinal cord injury), 파킨슨병(Parkinson's disease), 헌팅تون 병(Huntington's disease), 근위축 척삭증(amyotrophic lateral sclerosis), 연축성 발성장애(spasmodic dysphonia) 등을 앓고 있는 환아에게 어음향상 시스템을 통해 크고 명확한 소리 전달을 가능하게 하고, 전화기에 삽입되어 원활한 의사소통이 가능하게 하기 위한 장치이다. 이것은 음성에 장애를 갖고 있는 환아를 대상으로 개발된 시스템으로 임상적 유용성이 인정되고 있으나, 대부분 뇌신경계 질환에 의한 음성장애 환아들에 대해 적용되고 있다. 이 환아들은 비록 음성이 정상인과 다른 형태로 왜곡되어 발생하긴 하나, 음성을 발생하는 구조는 정상적인 형태로 존재하고 있는 경우를 대상으로 하고 있어 구개인두부전 환아와 같이 발성 구조가 손상되어 있는 경우에 대해서는 적용하기 어렵다. 또한 전체적인 시스템의 성능이 아직 미흡하여 음성의 증폭과 필터링에 의한 개선을 수행하고 있지만 환아의 개인에 맞추어 성능을 최적화시키는 기술은 불충분하여 환아에 적합한 최적의 시스템으로 성능을 나타내지 못한다. 그리고 영어권의 음성에 대하여 개발된 알고리즘을 사용하고 있기 때문에 한국어의 특성에 적합하지 못하다. 음성신호처리기술은 언어에 의존적이므로 기술도입을 하더라도 잡음제거와 같은 공통적인 문제에 대한 부분적인 도입만이 가능할 뿐, 한국어에 대해 본 과제를 성공

적으로 완수하기 위한 핵심기술을 개발한다는 측면에서 외국 기술도입은 불가능하다고 보여진다.

구개인두부전 환아의 한국어 코퍼스 구축은 한국어를 사용하는 정상인에 대한 음성신호처리 및 인간-기계 인터페이스 기술 발전에 기여하고, 음성인식을 이용한 내용 개발에 기여 할 뿐 아니라 구개인두장애 진단 및 계측장비 개발에 도움을 줄 수 있다. 물론 다른 장애의 음성 코퍼스 구축에도 응용이 가능하다. 그리고 구개인두부전 음성연구에서 정상인과 환아의 구강음/비강음을 분리하여 데이터를 구축함에 따라 음성 언어분야의 발전을 꾀하여 볼 수 있고, 학제 간 연구를 통한 공학과 의학의 발달 및 새로운 학문분야 개척 및 다양한 기술의 총망라로 연쇄적인 기술개발과 그에 따른 새로운 가치 산업을 창출하는 데 기본이 될 것이다. 음성장애인의 입장에서는 독립적인 생활 영위를 통한 삶의 질 개선과 사회간접비용과 가족들에 대한 의존도 감소, 장기적인 의료비 저축 감소, 의학적 치료자의 입장에서는 연령에 따라 증가하는 후천적 발성 장애에 대한 대비를 갖추는 기본 토대가 될 수 있을 것이다.

이러한 목표를 실현하기 위해서 우리 코퍼스 구축의 방법상 몇 가지 보완해야 할 사항이 있다. 본 연구에서는 음성녹음을 위한 발화 제시자료로서 기존 음성치료에 사용하는 문장과, 기존 연구를 통해 발음상의 오조음이 흔하게 나타난다고 알려진 모음, 자음, 어구들을 사용하였다. 구개인두부전 장애의 진단, 치료 및 재활을 위한 우리 코퍼스의 최종 사용 목적을 고려할 때 발화 문장으로써 역할을 가장 잘 해내는 문장의 선별작업이 필요할 것으로 생각된다. 이것은 정상인과 구개인두부전 환아의 자유발화 음성파일 비교 과정을 통하여 구개인두부전에 특징적인 문장, 어구의 발굴을 통해 가능할 것이라고 생각된다.

또한, 비강 구강 분리음을 위해 사용된 플라스틱 판의 음향

에너지 흡수여부의 확인과정을 거쳐야 하고, 어떤 재질의 분리판을 사용하는 것이 발화음 고유의 음향에너지에 미치는 영향을 최소화 할 수 있는지 다양한 재질을 사용하여 비교 검토하는 과정이 필요하다.

한편, 정상인에서 넬라톤 카테터로 경도, 중증도 구개인두부전 장애를 유발시켰을 때, 실제 구개인두부전 환아의 어느 정도의 장애상태와 동등한 관계에 있는지 확인할 수 있어야 한다. 본 연구에서 모의 장애환아에서 비음을 가진 문장을 발화/녹음하였을 때 예상 비음도보다 낮게 나왔는데, 이것은 우리 녹음 기기가 정확히 구음/비음 분리를 못하고 있거나, 넬라톤 카테터로 충분한 장애가 유발되지 않을 가능성을 시사한다. 따라서 넬라톤 카테터 위치를 다양하게 하여 모의 구개인두부전 환아와의 중증도 단계를 자유롭게 조작할 수 있고, 구개인두부전이 잘 형성되었는지 검증할 수 있도록 하는 방법을 보완해야 할 필요가 있다.

Acknowledgments

This study was supported by a grant from the Ministry of Health and Welfare (No. 2010-0020859).

REFERENCES

- 1) 김홍규, 강범모. 고려대학교 한국어 말모둠 1 (Korea-1 Corpus): 설계 및 구성. 서울, 한국: 한국어학;1996. p.233-258.
- 2) 권도하, 김동언, 김영태, 석동일, 정옥란. 구개파열 언어치료. 서울, 한국: 한국 언어치료학회;1995. p.85-93.
- 3) 김선희. 한국어 자유 발화 음성의 억양 패턴. 말소리와 음성과학 2009; 1(4):85-94.
- 4) Fletcher SG. Theory and instrumentation for quantitative measurement of nasality. Cleft Palate J 1970;7:601-9.
- 5) 양병곤. 프라트를 이용한 음성 분석의 이론과 실제. 만수출판사;2003. p.50-75.
- 6) 김진해. Nature of meaning: a perspective of corpus linguistics. 한국어의미학 2006;21:75-104.