Effects of Contextual Predictability Clues in Speech Materials on the Korean Speech Intelligibility Index

Hongyeop Oh¹, Soon-Je Choi¹, In-Ki Jin²

¹Department of Speech Pathology and Audiology, Graduate School, Hallym University, Chuncheon, Korea

²Division of Speech Pathology and Audiology, Research Institute of Audiology and Speech Pathology, College of Natural Sciences, Hallym University, Chuncheon, Korea

Received: March 17, 2020

Revised: April 30, 2020

Accepted: May 20, 2020

Correspondence:

In-Ki Jin, PhD

Division of Speech Pathology and Audiology, Research Institute of Audiology and Speech Pathology, College of Natural Sciences, Hallym University, 1 Hallymdaehak-gil, Chuncheon 24252, Korea

Tel: +82-33-248-2221

Fax: +82-33-256-3420

E-mail: inkijin@hallym.ac.kr

Purpose: This study aimed to derive band-importance functions (BIFs) and transfer functions (TFs) according to contextual predictability clues to determine the influence of contextual predictability clues in Korean speech material on the speech intelligibility index (SII). **Methods:** This study was conducted on 156 native speakers of Korean who had normal hearing. Korean speech perception in noise test material, which was composed of 120 high-predictability and 120 low-predictability sentences, was used for stimuli. To obtain intelligibility data, participants were tested for intelligibility in various frequency ranges and signal-to-noise ratio conditions. In order to derive the BIF and the TF, a nonlinear optimization procedure using MATLAB (MathWorks, Inc.) was used. **Results:** The BIF derived from the high-predictability sentences showed a peak in areas of 700 Hz (7.0%), 1,850 Hz (8.5%), and 4,800 Hz (7.6%). The crossover frequency for the high-predictability sentences was around 1,370 Hz. The BIF derived from the low-predictability sentences showed a peak in areas of 570 Hz (7.5%), 1,850 Hz (9.3%), and 4,000 Hz (8.0%). The crossover frequency for the low-predictability sentences was around 1,600 Hz. In the case of the TF, the TF curves derived from high-predictability sentences were steeper than those derived from low-predictability sentences. **Conclusion:** In the SII model, speech intelligibility differs according to contextual predictability clues. Especially, the more contextual predictability clues at the identical audibility, the higher the intelligibility predicted by the SII. Therefore, accurate speech intelligibility prediction requires the use of SII considering the contextual predictability clues that are characteristic of the stimulus.

Key Words: Band-importance function, Contextual predictability clues, Speech intelligibility index, Transfer function.

INTRODUCTION

어음명료지수(speech intelligibility index, SII)는 주어진 주파수 영역에서 가청도(audibility)를 계산하여 어음명료도의 수행력을 예측하는 모델이다(American National Standards Institute (ANSI), 1997; Amlani et al., 2002). 어음명료지수 산출에 있어 핵심 구성 요소로는 주파수중요함수(band-importance function, BIF)와 전달함수(transfer function)가 있다. 주파수중요함수는 음성을 청취하는 데 있어 전체 음성 정보에 대해 상대적으로 각 주파수 대역이 가지는 중요성을 수치로 표현한 함수를 의미하며, 전체 주파수에 대한 중요도를 1.0으로

설정하고 각 주파수 대역마다 상대적인 중요도를 할당하여 주파수별 중요도를 산정한다(Amlani et al., 2002; ANSI, 1997). 전달함수는 어음명료지수 값과 어음명료도 점수 간의 관계를 백분율로 나타낸 함수를 의미하며, 전달함수를 통해 어음명료지수 값으로부터 어음명료도의 수행력을 예측할 수 있다(Amlani et al., 2002; ANSI, 1997).

어음명료지수는 언어의 특성에 따라 달라질 수 있다. Wong et al.(2007)과 Jin et al.(2015)은 광동어와 한국어로 산출된 주파수중요함수가 영어로 산출된 주파수중요함수와 다른 특징을 가지는 것으로 보고하였다. 영어로 이루어진 주파수중요함수의 경우, 중주파수 대역에 밀집된 중요도를 나타내는 반면에 한국어와 광동어의 경우에는 저주파수 대역에서 높은 중요도를 보였다(Jin et al., 2015; Wong et al., 2007). 이는 어음명료지수 산출에 사용한 언어에 따라서 어음명료지수 값이 달라질

© 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.

수 있음을 시사한다.

어음명료지수는 또한 동일한 언어 내에서도 음성 자료의 특징에 따라 달라질 수 있다(ANSI, 1997). Studebaker et al.(1987)은 담화를 사용하여 주파수중요함수를 산출하였으며, Black(1959) 연구에서 산출한 단어에 따른 주파수중요함수와 비교하였다. 그 결과 단어로 산출한 주파수중요함수의 경우 두 개의 정점(peak)을 나타낸 반면에, 담화로 산출한 주파수중요함수는 넓은 대역에서 일정한 중요도를 보였다(Black, 1959; Studebaker et al., 1987). Yoho et al.(2018)은 화자의 특징과 어음 자료(speech material)에 따른 주파수중요함수 변화에 대해 보고하였다. 연구 결과에 따르면 자극음 녹음에 참여한 화자가 단일 화자와 다화자일 때의 주파수중요함수 차이를 구하였으며, 단일 화자에 비해 다화자일 때 주파수중요함수 중요도의 정점이 낮아지며 주변 주파수별 중요도 값의 차이가 줄어드는 것으로 나타났다. 또한 화자의 성별에 따라서 주파수중요함수 값의 차이를 비교하였을 때 남성 화자에 비해 여성 화자로 녹음된 자극음을 사용하였을 경우 전체적인 주파수중요함수 형태가 고주파수 영역으로 중요도의 비중이 이동한다고 보고하였다(Yoho et al., 2018). 또한 어음 자료 종류에 따른 차이를 비교하였을 때 단어와 문장의 주파수중요함수를 비교한 결과, 문장의 경우 1,600 Hz 영역에서 가장 높은 중요도를 나타낸 반면 단어의 경우 700 Hz 영역에서 가장 높은 중요도로 나타났다(Yoho et al., 2018). 선행연구 결과들은 어음 자료의 종류, 화자의 수, 화자의 성별과 같은 음성 자료의 특징이 주파수중요함수의 산출에 영향을 줄 수 있으며 이는 어음명료지수의 산출에도 영향을 줄 수 있음을 시사한다.

어음명료지수의 핵심 구성 요소인 주파수중요함수와 전달함수는 어음 자료의 특징인 문맥 예측 단서(contextual predictability clue)에 따라서도 달라질 수 있다. 첫째, 어음 자료의 문맥 예측 단서는 주파수중요함수의 산출에 영향을 줄 수 있다. DePaolis et al.(1996)은 단어, 문장 및 담화를 통해 주파수중요함수를 산출하였으며, 비교 결과로는 문맥 예측 단서가 낮은 자료일수록 가장 높은 중요도를 나타내는 정점이 상대적으로 높게 나타났으며 가장 높은 중요도 인근에 위치한 주파수 영역의 중요도가 낮게 나타났다. 예를 들어 단어의 경우 문장과 담화에 비해 정점이 두드러지게 높게 나타났으며, 정점 인근에 위치한 주파수 영역의 중요도는 상대적으로 감소하였다(DePaolis et al., 1996). Healy et al.(2013)은 영어 어음 자료를 사용하여 예측 단서가 높은 문장과 예측 단서가 낮은 문장의 주파수중요함수를 비교하였으며, 예측 단서가 낮은 문장의 주파수중요함수가 더 높은 정점을 나타낸다고 보고하였다. 둘째, 어음 자료의 문맥 예측 단서는 전달함수의 산출에도 영향을 미치는 것처럼 보인다. Amlani et al.(2002)은 문장, 단음절 및 무의미 음절

(nonsense syllable)의 전달함수를 산출하였으며, 비교 결과로는 문장, 단음절, 무의미 음절 순으로 전달함수의 기울기가 가파르게 나타났다. 자극음의 종류에 따른 전달함수의 차이에는 여러 원인이 존재할 수 있지만, 문맥 예측 단서가 그중 하나의 원인이 될 수 있다. 예를 들어 단음절 ‘감’은 목표 단어인 ‘감’을 제외하면 예측 단서가 존재하지 않는다. 반면에 ‘영화를 상영하는 장소는 극장입니다’라는 문장에서는 ‘영화’, ‘상영’, ‘장소’라는 표현은 목표 단어인 ‘극장’을 예측하는 데 있어 문맥적인 단서가 될 수 있다. 선행연구 결과들은 어음 자료의 특징인 문맥 예측 단서에 따라서 주파수중요함수 및 전달함수가 다르게 산출될 수 있는 가능성을 시사한다.

어음 자료의 문맥 예측 단서에 따라서 어음명료지수 값이 달라질 수 있기 때문에 문맥 예측 단서에 따른 어음명료지수 값을 산출하는 것은 정확한 어음명료도 수행력 예측에 있어 중요할 수 있다. Lindblom(1990)은 문장 내 문맥 예측 단서를 의미하는 문맥-상황적 정보와 음향-음소적 정보가 인간의 어음 인지 능력에 있어 중요한 영향을 미치는 것으로 보고하였다. 이는 문맥 예측 단서가 낮은 문장을 들었을 경우와 문맥 예측 단서가 높은 문장을 들었을 때에 어음 인지 수행력이 다를 수 있음을 시사한다. 문맥 예측 단서에 따른 어음명료지수의 영향에 대한 선행연구들이 있었지만 이러한 결과들은 주로 영어로 이루어진 자극음에 대한 결과들이었다(DePaolis et al., 1996; Healy et al., 2013; Studebaker et al., 1987). 하지만 영어와 다른 언어적 특성을 가지고 있는 한국어를 사용하여 문맥 예측 단서에 따른 어음명료지수 영향을 산출한 연구는 진행된 바가 없다. 만약 한국어를 기반으로 산출한 어음명료지수가 문맥 예측 단서에 따라 달라진다면 어음명료도 수행력을 예측할 때 사용하는 어음명료지수는 어음 자극의 문맥 예측 단서 특성을 고려한 적용이 필요함을 시사한다.

본 연구의 목적은 한국어 어음 자료의 문맥 예측 단서가 어음명료지수에 미치는 영향을 알아보고자 문맥 예측 단서가 높은 문장과 문맥 예측 단서가 낮은 문장이 구분되어 있는 Korean speech perception in noise (K-SPIN) (An et al., 2002; Kim et al., 2000) 어음 자료를 사용하여 주파수중요함수와 전달함수를 산출하고자 하였다. 본 연구의 가설은 다음과 같다. 첫째, 문맥 예측 단서에 따라 주파수중요함수의 주파수별 중요도는 달라질 것이다. 예를 들어 DePaolis et al.(1996)과 Healy et al.(2013)의 주파수중요함수 결과와 같이 문맥 예측 단서가 높을수록 주파수중요함수의 정점은 낮아지며 주변 주파수 영역의 중요도가 높아질 것이다. 그러나 영어의 문맥 예측 단서에 따른 주파수중요함수와는 다른 한국어 특성을 가진 주파수 중요도를 보일 것으로 예상된다. 둘째, 문맥 예측 단서에 따라 전달함수의 차이가 나타날 것이다. 예를 들어 문맥 예측 단서가 높을수

록 전달함수의 그래프 기울기는 가파르게 나타날 것이다. Am-lani et al.(2002)은 종류가 다른 어음 자료(문장, 단어, 무의미 단음절)를 통해 산출한 전달함수의 차이를 문맥 예측 단서가 높을수록 높은 어음명료도를 산출하는 것으로 보고하였다. 만약 문맥 예측 단서가 전달함수에 영향을 미친다면 동일 어음 자료 내에서도 문맥 예측 단서에 따른 차이가 나타날 것이다. 어음명료지수의 산출 공식에서 주파수중요함수와 전달함수는 핵심 구성 요소이다(ANSI, 1997). 따라서 문맥 예측 단서에 따라 주파수중요함수 또는 전달함수 값이 다르게 산출된다면 어음명료지수의 예측 값도 달라질 수 있다.

MATERIALS AND METHODS

연구 대상

본 연구는 한국어를 모국어로 사용하는 156명(남자 53명, 여자 103명)의 건청 성인을 대상으로 실시하였다(평균 연령: 21.6 세, 연령 범위: 만 19~28세). 모든 연구 참여자들은 과거에 이과 적 병력이 없었으며 순음청력검사 결과 250~8,000 Hz 영역에서 20 dB HL 이내의 정상 범위의 역치 값을 나타냈다. 고막운동도검사(tympanometry) 결과로는 A 유형으로 정상적인 고막 운동도를 나타냈다.

본 연구의 모든 연구 참여자에게는 연구를 시작하기 전 연구의 목적, 방법 및 진행 절차에 대해 충분한 설명을 제공하였으며, 연구 참여자들은 이에 대해 자발적으로 연구 대상자 동의서에 서명을 진행한 후 연구를 참여하였다. 또한 본 연구는 음성 녹음 파일 청취를 통해 들은 대로 컴퓨터에 작성하는 방식이 주요한 방법이기 때문에 연구 참여자가 연구 진행 중 휴식을 원하거나 피로감을 느낄 때 충분한 휴식 시간을 제공하였다. 연구가 종료된 후에는 모든 연구 참여자들에게 소정의 사례비를 지급하였다.

자극음

본 연구는 한국어 문장 내 문맥 예측 단서에 따른 어음명료지수에 차이를 구하고자 K-SPIN 어음 자료의 240문장을 사용하였다. K-SPIN 어음 자료의 구성으로는 문장당 2개에서 3개의 문맥 예측 단서를 포함하는 120개의 문맥 예측 단서가 높은 문장과 문맥 예측 단서를 포함하지 않는 120개의 문맥 예측 단서가 낮은 문장으로 구성되어 있다. 문맥 예측 단서가 높은 문장의 예로는 “여름이 더운 것은 이글거리는 태양 때문이다”이며, 이에 해당하는 질문으로는 “무엇 때문에 덥습니까?”이다. 문맥 예측 단서가 낮은 문장의 예로는 “이것은 생김새가 가위처럼 보인다”이며, 이에 해당하는 질문으로는 “무엇처럼 보입니까?”이다. K-SPIN 어음 자료는 여성 아나운서를 통해 이중벽

방음실(double-wall sound booth)에서 녹음하였으며 만들어진 K-SPIN 음원을 Adobe Audition 소프트웨어 버전 3.0 (Adobe Systems, San Jose, CA, USA)을 사용하여 표본화 주파수를 44,100 Hz로 양자화(quantization)는 16 비트(bit)로 설정하였다. 또한 어음명료지수는 다양한 소음 상황에서 측정된 어음명료도 수행력 점수를 통해 산출되기 때문에 본 연구의 음원들을 코딩소프트웨어(MATLAB version R2019a, MathWorks, Inc., Natick, MA, USA)를 통해 5가지(-8, -4, 0, +4, +8)의 신호대잡음비(signal-to-noise ratio, SNR)로 설정하였다. 5개의 신호대잡음비로 설정된 음원들은 21개의 저주파수 통과 필터(200, 300, 400, 510, 630, 770, 920, 1,080, 1,270, 1,480, 1,720, 2,000, 2,320, 2,700, 3,150, 3,700, 4,400, 5,300, 6,400, 7,700, 9,500 Hz)와 21개의 고주파수 통과 필터(100, 200, 300, 400, 510, 630, 770, 920, 1,080, 1,270, 1,480, 1,720, 2,000, 2,320, 2,700, 3,150, 3,700, 4,400, 5,300, 6,400, 7,700 Hz)를 통해 필터링하였으며 42개의 필터링 조건은 어음명료지수 임계 대역(SII critical band) 계산 절차를 따랐다(ANSI, 1997). 모든 필터의 기울기는 옥타브 단위당 96 dB로 설정하였다. 따라서 모든 자극음은 5개(-8, -4, 0, +4, +8 dB SNR)의 신호대잡음비와 42개의 필터(저주파수 통과 필터 21개, 고주파수 통과 필터 21개)를 사용하여 총 210가지 조건으로 구성하였다. 본 연구의 자극음 조건인 5개의 신호대잡음비는 본 연구와 유사한 연구 디자인과 연구 참여자 조건을 사용하여 안정적으로 한국어 기반 어음명료지수를 산출한 연구의 신호대잡음비 조건을 사용하였다(Heo et al., 2018; Jin et al., 2015).

실험 설계

일반적으로 사용되는 집단 간 설계와 집단 내 설계는 단독적으로 사용 시 통계적 검정력을 높일 수 있는 대표적인 실험 설계 방법으로 알려져 있다. 하지만 본 연구의 대상자 수와 사용되는 조건 수에 따라 필요한 문장 수의 한계로 인해 집단 간 설계(between-group design)와 집단 내 설계(within-subject design)를 결합한 혼합 설계(mixed design)를 사용하여 실험을 설계하였다. 집단 간 설계의 전제로는 각 그룹이 실험에서 서로 다른 조건에 참여해야 한다는 것이다. 따라서 본 연구의 경우 각 그룹은 총 210개의 조건 중 하나의 조건에만 참여할 수 있다. 하지만 본 연구에서 집단 간 설계를 사용할 경우 많은 대상자를 필요로 하기 때문에 집단 간 설계를 사용하지 못하였다. 예를 들어 각 그룹당 10명의 대상자가 필요한 경우 210가지의 조건을 모두 테스트하려면 총 2,100명의 대상자를 필요로 한다. 집단 내 설계의 전제로는 동일한 대상자가 모든 실험 조건에 참여해야 한다는 것이다. 따라서 본 연구의 경우 각 대상자는 총 210개 조건에 참여해야 한다. 하지만 본 연구에서 집단

내 설계를 사용할 경우 조건을 구상하기 위해 많은 문장을 필요로 하기 때문에 집단 내 설계를 사용하지 못하였다. 예를 들어 각 조건에 대해 10개의 문장이 고려되면 210개의 조건은 2,100개의 문장을 필요로 한다. 현재 국내에서 사용되는 어음 자료에는 한글로 구성된 2,100개의 문장을 포함하는 어음 자료는 없다.

따라서 필요로 하는 대상자 수와 총 210개 조건을 만족하는 어음 자료가 없는 한계로 인해 본 연구는 집단 간 설계와 집단 내 설계를 포함하는 혼합 설계를 사용하였다. 또한 본 연구에서는 실험 오차 값을 줄이고 완비블록 설계보다 필터링과 신호 대잡음비 조건을 정확하게 비교할 수 있는 균형불완비블록 설계(balanced incomplete block design)를 사용하였다. 완비블록 설계는 모든 실험 조건이 실험의 각 블록에 해당하는 것을 의미한다. 하지만 실험 조건의 수가 블록의 수를 초과하면 균형불완비블록 설계를 사용해야 한다(Hinkelmann & Kempthorne, 2005). Yates(1936)에 의해 제안된 균형불완비블록 설계는 Fleiss(1981), Jung(1961), Yang et al.(2011) 등의 연구에서 사용하였다. 혼합블록 설계를 사용하기 위해서는 공식 [1]과 같이 대상자 수와 자극을 수를 곱한 값이 반복 횟수와 총 조건의 수를 곱한 값과 같아야 한다.

$$\begin{array}{ccccccc} \text{Number} & \text{Number} & \text{Number} & \text{Number} & \text{Total} \\ \text{of} & \text{of} & \text{of} & \text{of} & \text{number} \\ \text{subjects} & \text{stimuli} & \text{repetitions} & \text{conditions} & \text{of data} \end{array} \quad [1]$$

$$156 \times 24 = 18 \times 208 = 3,744$$

본 연구는 공식 [1]의 조건을 충족하기 위해 210개의 조건에서 0%의 어음명료도로 예상되는 -8 dB 신호대잡음비 조건에서 저주파수 통과 필터를 사용한 중심 주파수 기준 200 Hz의 조건과 고주파수 통과 필터를 사용한 중심 주파수 기준 7,500 Hz의 조건을 제외하여 총 208개의 조건을 본 연구의 혼합블록 설계로 선택하였다. 따라서 어음명료도 측정 시 대상자들에게 208개의 조건 중 24개의 조건에 해당하는 문장들을 무작위로 제시하였다. 위와 같은 실험 설계로 인하여 연구 대상자들은 중복된 문장을 듣지 않았고 각 조건당 18개의 데이터 수집이 가능하였다.

본 연구와 유사한 방법으로 진행한 선행연구(Heo et al., 2018; Jin et al., 2015; Oh et al., 2019)들은 공식 [1]과 유사한 혼합블록 설계를 사용하였다. 선행연구들은 대상자 수(78) × 자극 수(24) = 반복 횟수(9) × 총 조건의 수(208) = 총 데이터 수(1,872)와 같은 혼합블록 설계식을 사용하였으며 본 연구에 비해 대상자 수와 반복 횟수의 수는 1/2의 값에 해당하였다. 본 연구와 선행연구들에 혼합블록 설계가 다르게 구성된 이유는 다음과 같다. 첫째, 본 연구와 선행연구의 어음명료도 측정에

사용하는 문장의 수가 달랐다. 본래의 K-SPIN 문장의 경우 1개의 리스트당 40개의 문장을 포함하며 총 6개의 리스트(240문장)로 구성되어 있다. 하지만 본 연구와 선행연구들은 실험 설계 조건을 만족하기 위해 1개의 리스트당 10개의 문장을 포함하는 총 24개의 리스트로 재구성하여 사용하였다. 재구성한 리스트들의 문장 구성 비율(문맥 예측 단서가 낮은 문장:문맥 예측 단서가 높은 문장)은 최대한 동일하게 유지하였다. 기존의 선행연구들은 재구성된 리스트를 통해 하나의 검사 조건당 10개의 문장을 사용하여 어음명료도를 측정하였다. 하지만 본 연구는 문맥 예측 단서에 따른 어음명료도를 측정하기 위해 문맥 예측 단서가 높은 문장에 대한 데이터와 문맥 예측 단서가 낮은 문장에 대한 데이터를 분리하여 어음명료도를 산출하였다. 이는 선행연구와 동일한 실험 설계를 사용할 시 한 조건에 누적되는 데이터의 양이 선행연구의 1/2에 해당하는 것을 의미하기도 한다. 따라서 선행연구 대상자 수의 두 배에 해당하는 대상자를 모집하여 한 조건에 누적되는 데이터의 양을 보완하였다.

어음명료도 측정 절차

본 연구의 어음명료도는 선행연구의 방법과 동일한 절차로 진행하였다(Heo et al., 2018; Jin et al., 2015). 학습 효과를 막기 위해 각 연구 참여자는 총 420개 조건(42 필터링 × 5 신호 대잡음비 × 2 문맥 예측 단서 = 420개 조건) 중 24개 조건(각 조건당 10문장)에만 무작위 순서로 참여하였다. 따라서 각 참여자들은 동일 문장을 1회 이상 청취하지 않았으며 각 조건당 수집한 데이터는 18개였다. 모든 연구 참여자들은 컴퓨터에서 제시되는 자극음을 GSI-61 (Grason-Stadler Inc., Eden Prairie, MN, USA)과 연결된 TDH-50 (Telephonics, Farmingdale, NY, USA) 헤드폰을 통해 오른쪽 귀로 청취하였다. 연구 참여자들에게는 음원과 음원에 대한 질문을 듣고 답변을 컴퓨터에 받아쓰도록 안내하였다. 예를 들어 연구 참여자는 “영화를 상영하는 장소는 극장입니다”와 같은 음원을 듣고 “어디입니까?”와 같은 질문 음원을 통해 정답(극장)의 정확한 기재 여부를 확인하였다. 또한 연구를 시작하기 전 자체 녹음한 SPIN 문장과 관계없는 문장 5개(예: ‘어제 영화가 먹은 음식은 족발이었다’)를 사용하여 연습 기회를 제공하였으며, 연구 참여자들이 제시한 문장 전체가 아닌 일부만을 들었을 경우에도 추측하여 최대한 질문에 대한 답변을 작성하도록 권고하였다. 한 조건에 대한 어음명료도는 10개의 문장에서 핵심 단어(keyword)를 정확히 맞춘 개수를 백분율로 측정하였다. 한 명의 연구 참여자에게 할당된 검사 조건에 대한 어음명료도를 측정하는 소요 시간은 약 1시간 30분 정도였으며 참여자들이 원할 경우 언제든지 휴식을 취할 수 있도록 안내하였다. 모든 어음명료도 측정 절차는 ANSI/ASA S3.1-1999 (ANSI, 2013)의 주변 소음 기준을 만

족하는 이중벽 방음실 내에서 이루어졌다.

어음명료지수 산출 방법

어음명료지수는 공식 [2]에 따라 산출하였다. F 는 주파수 대역(frequency band)의 전체적인 숫자를 의미하고 BIF_f 는 주파수 대역 k 에서의 주파수중요함수(BIF)를 의미하고 BAF_f 는 주파수 대역 f 에서의 대역가청함수(band-audibility function, BAF)를 의미한다(ANSI, 1997). 따라서 어음명료지수는 주파수 대역별로 대역가청함수와 주파수중요함수를 곱하여 각 주파수 대역별 어음명료지수 값을 산출하고 각 주파수 대역별로 산출한 어음명료지수 값을 합하여 총 어음명료지수 값을 산출하였다. 어음명료지수 값은 0에서부터 1 사이의 값으로 표현되며 수치가 1에 가까울수록 높은 어음명료도를 뜻한다.

$$SII = \sum_{f=1}^F BIF_f \times BAF_f \quad [2]$$

어음명료지수의 어음명료도 예측은 어음명료지수 값과 어음명료도의 백분율 점수 사이의 관계를 나타내는 전달함수를 통해 산출할 수 있다. 전달함수는 공식 [2]에 따라 산출하였다.

‘S’는 어음명료도 값을 의미하고 ‘P’는 청자와 화자의 숙련도(숙련도 값의 범위는 0~1이며 0은 최저의 숙련도를, 1은 최고의 숙련도를 의미한다)를 의미하며 ‘A’는 공식 [3]에서 계산된 어음명료지수 값을 의미한다. ‘Q’와 ‘N’은 함수 곡선의 왜곡을 최소화하여 산출하기 위한 상수(fitting constant) 값을 의미한다.

$$S = (1 - 10^{\frac{PA}{Q}})^N \quad [3]$$

본 연구에서는 앞서 언급한 산술식들을 근거하여 MATLAB 최적화 도구에서 제공하는 Kates(2013)가 개발한 비선형최적화함수(nonlinear optimization function)를 사용하였다. 비선형최적화함수는 데이터 평활화(data smoothing)와 그래픽 설계(graphical construction) 과정을 거치지 않은 명료도 데이터(raw data)를 사용하여 주파수중요함수와 전달함수를 도출할 수 있다. 비선형최적화함수의 최적화 기준은 공식 [2]의 평균 제곱근(root mean square) 오차를 최소화하며 검사 조건 전체에서 확인된 어음명료도의 평균 비율을 통합하는 방법을 사용한다. 또한 Kates(2013)는 비선형최적화함수의 수학적 절차로 주파수중요함수와 ‘Q’와 ‘N’의 값을 동시에 조정하여 어음명료지수가 연구 참여자의 명료도 점수에 최적화되도록 제작하였다.

MATLAB 프로그램을 사용하여 오류를 최소화하는 방법으로는 다음과 같이 진행하였다. 첫째, 매개변수 ‘Q’ 및 ‘N’ 값과 함께 주파수중요함수의 초기 추정치를 포함하는 벡터로 설정하였으며 역동 범위는 30 dB로 설정하였다. 또한 초기 매개 변수는

‘Q’ 값 0.2178과 ‘N’ 값 15.52로 설정하였으며 21개의 주파수중요함수 가중치는 모두 1/21로 초기화하였다. 둘째, MATLAB의 ‘fminsearch’ 함수를 사용하여 초기 추정치보다 오류를 줄인 주파수중요함수를 도출하였다. 셋째, MATLAB의 ‘fmincon’ 함수를 사용하여 한 번 더 오류 최소화 과정을 거쳤다(Lagarias et al., 1998). 이 방법을 사용하려면 내점(interior point)을 찾는 설정 값이 주파수중요함수의 경우에 합이 1이어야 하며 모든 매개변수와 각 주파수 대역별 중요도 값은 0 이상이어야 한다(Byrd et al., 1999). 마지막으로 5점 평활화(five-point smoother) 과정을 통해 주파수중요함수 결과의 이상치(outlier)를 제거하였다(Marchand & Marmet, 1983).

RESULTS

주파수중요함수 산출

문맥 예측 단서에 따른 주파수중요함수는 Figure 1에 제시하였다. 전체적인 주파수중요함수의 형태로는 3개의 정점이 나타났다. 문맥 예측 단서가 높은 문장의 주파수중요함수 결과는 첫 번째 정점은 중심 주파수 기준 700 Hz (7.0%) 영역에서 나타났고, 두 번째 정점은 중심 주파수 기준 1,850 Hz (8.5%) 영역에서 나타났으며, 세 번째 정점은 중심 주파수 기준 4,800 Hz (7.6%) 영역에서 나타났다. 교차 주파수는 중심 주파수 기준 1,370 Hz 영역에 형성되었고 1,370 Hz 이하는 중요도의 48.5%에 해당하였다. 문맥 예측 단서가 낮은 문장의 주파수중요함수 결과는 첫 번째 정점이 중심 주파수 기준 570 Hz (7.5%) 영역에서 나타났고, 두 번째 정점은 중심 주파수 기준 1,850 Hz (9.3%) 영역에서 나타났으며, 세 번째 정점은 4,000 Hz (8.0%) 영역에서 나타났다. 교차 주파수(crossover frequency)는 중심 주파수 기준 1,600 Hz 영역에 형성되었고 1,600 Hz 이하의 중요도는 49.7%에 해당하였다. 문맥 예측 단서가 높은 문장과 문맥 예측 단서가 낮은 문장의 주파수중요함수 차이는 0.1%(2,150 Hz)에서 3.8%(3,400 Hz)로 유사한 주파수중요함수 형태를 나타냈다. 예를 들어 1% 미만의 차이를 나타낸 중심 주파수는 150, 250, 1,600, 1,850, 2,150, 2,500, 7,000 Hz 영역에 해당하였으며 1% 이상 2% 미만의 차이를 나타낸 중심 주파수는 350, 570, 1,170, 1,370, 4,000, 8,500 Hz 영역에 해당하였으며 2% 이상 3% 미만의 차이를 나타낸 중심 주파수는 450, 700, 1,000, 2,900, 4,800, 5,800 Hz 영역에 해당하였다. 3% 이상의 차이를 나타낸 중심 주파수는 840, 3,400 Hz 영역이었다.

전달함수의 산출 및 측정과 예측 명료도의 상관성

문맥 예측 단서에 따른 K-SPIN 어음 자료의 전달함수는 Figure 2에 제시하였다. 문맥 예측 단서가 높은 문장으로 도출

Figure 1. Band-importance functions for high-predictability (circle) and low-predictability (triangle) as a function of 21 frequency bands.

Figure 2. Transfer functions for high-predictability stimuli (solid line) and low-predictability stimuli (dashed line). The X axis shows the SII value and the Y axis shows the percent of correct keywords. SII: speech intelligibility index.

된 전달함수 곡선이 문맥 예측 단서가 낮은 문장으로 도출된 전달함수 곡선보다 더 가파르게 나타났다. 예를 들어 문맥 예측 단서가 높은 문장의 0.3 SII 값은 80.7%의 어음명료도에 해당하는 반면, 문맥 예측 단서가 낮은 문장의 0.3 SII 값은 46.9%의 어음명료도에 해당하였다. 문맥 예측 단서가 높은 문장의 0.8 SII 값은 99.9%의 어음명료도에 해당하는 반면, 문맥 예측 단서가 낮은 문장의 0.8 SII 값은 98.7%의 어음명료도에 해당하였다.

본 연구에서 사용한 피팅 상수 값, 피어슨 상관관계수 값, 제공

Table 1. Best-fit values for fitting constants of Q, N, Pearson Corr, and RMS error for transfer function of high- and low-predictability stimuli

Stimuli	Q	N	Corr	RMS error
High-predictability	0.210	5.643	0.912	0.161
Low-predictability	0.287	8.006	0.931	0.127

Corr: correlation coefficient, RMS: root mean square

평균제공근 오차의 최적 값은 Table 1에 제시하였다. 피팅 상수 값(Q, N)은 주파수중요함수와 전달함수 그래프를 신뢰도 있게 산출하기 위해 조절된 상수 값을 의미하고 피어슨 상관관계수는 본 연구에서 산출된 어음명료지수를 통해 예측한 어음명료도와 연구 참여자에게서 실제로 측정된 어음명료도 점수 간의 상관관계를 의미하며, 제공 평균 제공근은 본 연구에서 산출된 어음명료지수를 통해 예측한 어음명료도와 연구 참여자에게서 실제로 측정된 어음명료도 점수 간의 오차 값을 의미한다. 따라서 문맥 예측 단서가 높은 문장의 경우 'Q'와 'N' 값은 각각 0.210, 5.643이었으며 예측된 어음명료도와 실제 측정된 어음명료도 간의 상관관계는 0.912, 오차는 0.161로 상관관계가 매우 높게 나타났다. 문맥 예측 단서가 낮은 문장의 경우 'Q'와 'N' 값은 각각 0.287, 8.006이었으며 예측된 어음명료도와 실제 측정된 어음명료도 간의 상관관계는 0.931, 오차는 0.127로 상관관계가 매우 높게 나타났다. 이는 본 연구의 어음명료지수를 통해 산출한 문맥 예측 단서에 따른 어음명료도와 실제로 측정된 어음명료도를 비교하였을 때 본 연구의 어음명료지수가 높은 예측 정확도를 나타낸다는 것을 의미한다.

DISCUSSIONS

본 연구는 K-SPIN 어음 자료를 사용하여 도출한 주파수중요함수와 전달함수를 통해 문맥 예측 단서가 어음명료지수에 미치는 영향을 확인하고자 하였다. 문맥 예측 단서가 높은 문장으로 산출한 주파수중요함수는 중심 주파수 기준 700 Hz (7.0%), 1,850 Hz (8.5%), 4,800 Hz (7.6%) 영역에서 정점을 나타냈으며 교차 주파수는 1,370 Hz 영역으로 나타났다. 문맥 예측 단서가 낮은 문장으로 산출한 주파수중요함수는 중심 주파수

기준 570 Hz (7.5%), 1,850 Hz (9.3%), 4,000 Hz (8.0%) 영역에서 정점을 나타냈으며 교차 주파수는 1,600 Hz 영역으로 나타났다. 문맥 예측 단서가 높은 문장과 낮은 문장의 주파수중요함수 차이는 2개(840, 3,400 Hz)의 중심 주파수를 제외한 19개의 중심 주파수 영역에서 3% 미만 차이를 보여 유사한 형태의 주파수중요함수를 갖는 것으로 나타났다. 본 연구의 결과는 영어 어음 자료의 문맥 예측 단서에 따른 어음명료지수의 영향을 보고한 선행연구들과 유사하게 문맥 예측 단서가 높을수록 주파수중요함수의 정점은 낮아지며 전달함수를 통해 예측되는

Figure 3. Phoneme distributions of keywords in Korean speech perception in noise sentences on the band-importance functions for high- (A) and low-predictability (B) stimuli as a function of 21 frequency bands. Each phoneme is located in an averaged frequency (Hz) area according to production. Numbers in parentheses represent the frequency of use for each phoneme and an unreleased stop represents a coda.

어음명료도의 수행력은 높게 나타났다(DePaolis et al., 1996; Healy et al., 2013). 또한 영어로 산출한 주파수중요함수와는 다르게 저주파수 영역에서 높은 정점을 나타냈다. 이는 영어에 비해 모음의 산출 빈도가 높아 저주파수 영역에서 높은 정점을 나타내는 한국어 주파수중요함수의 보편적인 특징이 반영된 것으로 보인다(Heo et al., 2018; Jin et al., 2015).

본 연구의 주파수중요함수 결과는 문맥 예측 단서에 따른 차이를 나타냈지만 선행연구 결과와 동일하게 주파수중요함수의 형태를 나타내는 정점과 골(valley)의 위치가 유사하게 나타났다(Healy et al., 2013). 문맥 예측 단서에 따른 주파수중요함수의 결과가 유사하게 나타난 것을 설명하기 위해 본 연구는 음성 전사 방법을 통해 K-SPIN 어음 자료의 정답에 해당하는 단어들을 분석하였다. 음소의 주파수 영역은 Kim et al.(2013)을 참고하였으며, 주파수 영역에 따른 초성, 중성, 종성의 개수와 주파수중요함수를 함께 나타낸 분포도를 Figure 3에 제시하였다. Figure 3을 보면 문맥 예측 단서가 높은 문장과 문맥 예측 단서가 낮은 문장의 초성, 중성, 종성의 빈도 차이가 적은 것을 알 수 있다. 예를 들어 문맥 예측 단서가 높은 문장들의 초성 [ㄴ]의 빈도 13개, 문맥 예측 단서가 낮은 문장들의 초성 [ㄴ]의 빈도 13개로 동일하였으며, 문맥 예측 단서가 높은 문장들의 중성 [ㄹ]의 빈도는 33개, 문맥 예측 단서가 낮은 문장들의 중성 [ㄹ]의 빈도는 34개로 유사하였다. 이는 문맥 예측 단서에 따라 주파수중요함수의 차이가 나타났음에도 정점과 골의 위치가 유사하게 나타난 것을 일부 설명한다. 또한 Figure 3을 보면 K-SPIN 어음 자료 정답에 해당하는 음소가 본 연구의 두 번째 정점, 세 번째 정점, 첫 번째 정점 순으로 많이 분포되어 있는 것을 확인할 수 있다. 주파수중요함수의 경우 다양한 조건(언어, 화자의 성별, 어음 자료의 종류)에 따라 주파수별 중요도가 달라질 수 있기 때문에 음소의 주파수 영역과 빈도수만으로는 본 연구의 결과를 전부 설명할 수 없다. 하지만 음소의 주파수 영역과 빈도수에 따라서 중주파수 영역에 가장 높은 정점과 한국어 주파수중요함수의 특징인 저주파수 영역의 높은 정점이 나타난 것을 확인할 수 있다. 이는 어음명료도 측정 시에 사용한 음소의 주파수 영역과 빈도수가 주파수중요함수 결과에 영향을 미칠 수 있는 것을 나타내며 추후 어음명료지수 산출에 사용한 음소 차이에 따른 주파수중요함수의 영향에 대한 연구가 필요함을 시사한다.

본 연구는 문맥 예측 단서 특성에 따른 어음명료지수의 차이를 확인하고자 하였지만 문장 특성에 따른 한계점도 다소 존재하였다. 문맥 예측 단서에 따른 차이를 더욱 명확하게 확인하기 위해서는 동일 문장에서 핵심 단어만을 바꾸어 어음명료도를 측정하는 것이 현재 적용된 방법보다 더욱 효과적으로 문맥 단서에 따른 차이를 명확히 확인할 수 있었을 것이다. 예를 들어

영화를 상영하는 장소는 극장입니다”라는 문장과 “영화를 상영하는 장소는 시장입니다”와 같이 문맥적 단서에 따른 핵심 단어만을 바꾸어 측정하였다면 문맥적 단서에 따른 차이를 더욱 명확히 확인할 가능성이 높았을 것이다. 본 연구의 결과가 문맥적 단서에 따른 어음명료지수의 차이를 확인하기 위한 일부 결과를 보여주었지만, 추후 문맥적 단서에 따른 더욱 명확한 차이를 확인하기 위해서는 추가적인 연구가 더 필요하다고 생각한다.

본 연구는 문맥 예측 단서가 어음명료지수의 핵심 구성 요소인 주파수중요함수와 전달함수에 영향을 미치는지 확인하고자 하였다. 그 결과 문맥 예측 단서에 따른 주파수중요함수는 0.1%에서 3.8%의 차이를 보였지만 2개의 중심 주파수를 제외하고는 3% 이하로 문맥 예측 단서가 다른 문장이라 할지라도 유사한 주파수중요함수의 형태를 나타내는 것을 확인하였다. 전달함수 결과로는 문맥 예측 단서에 따라 어음명료지수로 예측된 어음명료도 값이 달라지는 것을 확인하였다. 이러한 결과는 어음명료지수 모델이 문맥 예측 단서에 따라 어음명료도 수행력의 차이를 가져올 수 있다는 것을 의미하고 이는 보청기 적합 과정 시 문맥 예측 단서가 높은 어음 자료를 활용하여 보청기 적합 예측을 예측할 때와 문맥 예측 단서가 낮은 어음 자료를 활용하여 보청기 적합 예측을 예측할 때 보청기 착용자의 어음명료도에 차이가 나타날 수 있음을 의미한다. 본 연구 결과는 어음명료지수의 구성 요소인 주파수중요함수와 전달함수 값이 동일 언어 내에서 문맥적 특성에 따라 달라져 어음명료도 예측에 영향을 줄 수 있음을 보여주었다. 따라서 어음명료지수의 예측 값과 실제 측정된 어음명료도를 비교할 때, 어음명료지수와 실제 어음명료도 측정에서 사용된 어음 자극의 특성을 고려한 해석이 요구될 것이다.

중심 단어 : 주파수중요함수 · 문맥 예측 단서 · 어음명료지수 · 전달함수.

Ethical Statement

The study was approved by the Institutional Review Board of Halllym University (HIRB-2019-043).

Acknowledgments

The authors thank to Sanghun Park for creating a sound presentation program.

Declaration of Conflicting Interests

There are no conflict of interests.

Funding

This study was supported by Basic Science Research Program Grant NRF-2015R1C1A1A01052458 funded by the Ministry of Science, ICT and Future Planning.

Author Contributions

H.O. performed experiments, analyzed data, and wrote the paper; S.C.

performed experiments and analyzed data; I.J. designed experiments, analyzed data, and wrote the paper. Also, the authors discussed the results together and implications and commented on the manuscript at each stage.

ORCID iDs

Hongyeop Oh <https://orcid.org/0000-0002-5493-4648>
In-Ki Jin <https://orcid.org/0000-0002-0834-5981>

REFERENCES

- American National Standards Institute. (1997). *ANSI S3.5-1997 (R2012). American National Standard Methods for Calculation of the Speech Intelligibility Index*. New York, NY: Acoustical Society of America.
- American National Standards Institute. (2013). *ANSI/ASA S3.1-1999 (R2013). Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms*. New York, NY: Acoustical Society of America.
- Amlani, A. M., Punch, J. L., & Ching, T. Y. (2002). Methods and applications of the audibility index in hearing aid selection and fitting. *Trends in Amplification*, 6(3), 81-129.
- An, L. J., Kim, J. S., & Pae, S. Y. (2002). The study on developing a test of speech perception in noise. *Journal of Audiology and Otology*, 6(2), 118-127.
- Black, J. W. (1959). Equally contributing frequency bands in intelligibility testing. *Journal of Speech and Hearing Research*, 2(1), 81-83.
- Byrd, R. H., Hribar, M. E., & Nocedal, J. (1999). An interior point algorithm for large-scale nonlinear programming. *SIAM Journal on Optimization*, 9(4), 877-900.
- DePaolis, R. A., Janota, C. P., & Frank, T. (1996). Frequency importance functions for words, sentences, and continuous discourse. *Journal of Speech and Hearing Research*, 39(4), 714-723.
- Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement*, 5(1), 105-112.
- Healy, E. W., Yoho, S. E., & Apoux, F. (2013). Band importance for sentences and words reexamined. *The Journal of the Acoustical Society of America*, 133(1), 463-473.
- Heo, S., Oh, H., & Jin, I. K. (2018). Comparisons of the Korean band-importance and absolute transfer functions according to speakers' gender. *Audiology and Speech Research*, 14(4), 236-242.
- Hinkelmann, K. & Kempthorne, O. (2005). *Design and Analysis of Experiments. Volume 2, Advanced Experimental Design*. New York, NY: Wiley-Interscience.
- Jin, I. K., Kates, J. M., Lee, K., & Arehart, K. H. (2015). Derivations of the band-importance function: A cross-procedure comparison. *The Journal of the Acoustical Society of America*, 138(2), 938-941.
- Jung, A. F. (1961). Interviewer differences among automobile purchasers. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 10(2), 93-97.
- Kates, J. M. (2013). Improved estimation of frequency importance functions. *The Journal of the Acoustical Society of America*, 134(5), EL459-EL464.
- Kim, J. S., Pae, S. Y., & Lee, J. H. (2000). Development of a test of Korean speech intelligibility in noise (KSPIN) using sentence materials with controlled word predictability. *Speech Sciences*, 7(2), 37-50.
- Kim, J. S., Shin, E. Y., & Cho, E. B. (2013). A study on initial-middle-final phoneme frequency analyses of the Korean meaningful monosyllabic words. *Audiology*, 9(2), 127-136.
- Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1), 112-147.
- Lindblom, B. (1990). Models of phonetic variation and selection. In Lindblom, B. *Phonetic Experimental Research at the Institute of Linguistics* (pp. 65-100). Stockholm: University of Stockholm.
- Marchand, P. & Marmet, L. (1983). Binomial smoothing filter: A way to avoid some pitfalls of least-squares polynomial smoothing. *Review of Scientific Instruments*, 54(8), 1034-1041.
- Oh, H., Heo, S., & Jin, I. K. (2019). A comparison of band-importance function and transfer function using clear speech and conversational speech. *Audiology and Speech Research*, 15(3), 168-175.
- Studebaker, G. A., Pavlovic, C. V., & Sherbecoe, R. L. (1987). A frequency importance function for continuous discourse. *The Journal of the Acoustical Society of America*, 81(4), 1130-1138.
- Wong, L. L. N., Ho, A. H. S., Chua, E. W. W., & Soli, S. D. (2007). Development of the Cantonese speech intelligibility index. *The Journal of the Acoustical Society of America*, 121(4), 2350-2361.
- Yang, Z. F., Lee, P. C., Chen, W. H., & Leu, J. G. (2011). Extension of structural watermarks based on balanced incomplete block designs. *Journal of Information Hiding and Multimedia Signal Processing*, 2(4), 354-365.
- Yates, F. (1936). Incomplete randomized blocks. *Annals of Eugenics*, 7(2), 121-140.
- Yoho, S. E., Healy, E. W., Youngdahl, C. L., Barrett, T. S., & Apoux, F. (2018). Speech-material and talker effects in speech band importance. *The Journal of the Acoustical Society of America*, 143(3), 1417-1426.