



Reliability and Validity of the Korean Matrix Sentence-in-Noise Recognition Test for Older Listeners with Normal Hearing and with Hearing Impairment

Yukyung Jung¹, Jaeho Han¹, Hyo Jung Choi¹, Jae Hee Lee^{1,2}

¹Department of Audiology and Speech-Language Pathology, Hallym University of Graduate Studies, Seoul, Korea

²HUGS Center for Hearing and Speech Research, Seoul, Korea

Received: September 17, 2022

Revised: October 5, 2022

Accepted: October 13, 2022

Correspondence:

Jae Hee Lee, PhD

Department of Audiology and Speech-Language Pathology, Hallym University of Graduate Studies, 427 Yeoksam-ro, Gangnam-gu, Seoul 06197, Korea

Tel: +82-2-2051-4952

Fax: +82-2-3451-6618

E-mail: leejaehee@hallym.ac.kr

Purpose: The aim of this study was to examine test-retest reliability and validity of the adaptive Korean matrix sentence-in-noise test for elderly group with normal hearing (ENH) and elderly group with hearing impairment (EHI). **Methods:** Twenty ENH listeners and 20 EHI listeners based on the World Health Organization (WHO)-proposed hearing-impairment grade participated. For the test-retest reliability of the Korean matrix sentence-in-noise test, each subject was assessed twice in a 2-week interval using three types of noise: speech-shaped noise, International Collegium of Rehabilitative Audiology noise, and International Speech Test Signal. The test-retest agreement was judged by the intraclass correlation coefficient (ICC) and Bland-Altman plots. To differentiate between real change and random measurement error, the standard error of measurement (SEM) and the minimal detectable change (MDC) were also obtained as another reliability index. For the concurrent validity, the results of the Matrix test were compared to the results of the Korean-version Hearing In Noise Test (K-HINT). **Results:** The test-retest reliability was good to excellent for ENH and EHI groups (ICC, 0.84~0.98). Both groups showed stable reliability, but the ENH group showed smaller SEM and MDC values than the EHI group. No systematic errors were found from the Bland-Altman plots. The correlation between the results of the Matrix test and the K-HINT test was high, revealing good concurrent validity. **Conclusion:** The Korean matrix sentence-in-noise test appeared to be valid and reliable for the elderly. The Korean matrix sentence-in-noise test can be utilized for clinical measurement of the speech-in-noise ability of elderly individuals.

Key Words: Reliability, Validity, Korean matrix sentence test, Speech-in-noise intelligibility.

INTRODUCTION

난청은 관절염 및 고혈압과 함께 노년기에 흔히 발생하는 3대 만성 질환 중 하나로 알려져 있다. Lin et al.(2011) 보고에 따르면 60~69세 미국 인구의 약 45%, 70~79세 인구의 약 68%, 80세 이상의 노인의 경우 약 89%가 난청을 보이고 있다. 노인성 난청이 발생한 경우 고주파수에서 청력 역치가 상승하는 것뿐 아니라 배경소음 혹은 반향이 있는 상황에서 어음인지가 저하되어 노인의 의사소통 및 대인관계, 여가활동, 사회 경제적 활동 등 다양한 일상생활 영역에 영향을 미치고 주관적 삶의 질이 저하될 수 있다(Anderson & Kraus, 2010; Chang et al., 2009; Ciorba

et al., 2012; Dayasiri et al., 2011; Kim et al., 2021).

임상에서 난청인의 어려움을 확인하기 위해 순음청력역치 혹은 소음없는 조건에서 어음인지도를 측정하나, 이와 같은 기본청력검사 결과만으로 난청인의 소음 하 의사소통 능력을 유의하게 예측하기 어렵다(Kim & Lee, 2020; Taylor, 2003; Vermiglio et al., 2012; Wilson, 2011). 미국, 독일 등의 여러 국가들의 경우 난청인 대상자의 소음 상황 속 의사소통 능력을 측정하기 위해 Hearing in Noise Test (HINT; Nilsson et al., 1994), Quick Speech in Noise Test (Killion et al., 2004), Matrix test (Kollmeier et al., 2015) 등의 소음 하 어음인지검사 도구를 개발 및 표준화하여 진단 및 연구 등에 활용하고 있다.

국내의 경우 한국어 HINT (Moon et al., 2008)와 한국어 Matrix 검사(Kim & Lee, 2018; Kim & Lee, 2020; Jung et

al., 2021)가 변동형 소음 하 어음청각검사에 활용 가능하다. 두 검사 모두 대상자가 각 문장을 듣고 정반응을 보이면 문장 제시 레벨이 감소하고 오반응을 보이면 문장 제시레벨이 증가하는 등 대상자의 반응에 따라 문장 레벨이 자동 변동하게 된다. HINT와 Matrix 검사 모두 약 65 dB sound pressure level (SPL)의 고정된 레벨에서 소음을 제시한 채 청자가 문장을 50% 가량 인지할 수 있는 신호대잡음비(signal-to-noise ratio, SNR)를 검사 결과로 산출하는 점에서 유사하다. HINT 검사는 일상 회화체 문장을 목표 문장으로 제시하며, 검사 시 헤드폰뿐 아니라 정면(0°)과 좌우(± 90°)에 위치한 음장스피커를 활용하여 검사가 가능하다. HINT 검사는 Bio-logic 사의 HINT Pro 장비를 사용해야 하며, 현재 한국어 버전의 경우 검사 도구의 상용화 중단 문제로 임상적 한계점이 있다.

Matrix 검사는 10 × 5 행렬(matrix)에 포함된 50개 단어의 무작위 조합으로 이루어진 문장을 목표 문장으로 제시하므로 각 문장마다 동일한 구조를 가지며, HINT 검사와 마찬가지로 헤드폰뿐 아니라 음장스피커를 활용한 음장 어음청각검사가 가능하다. 한국어 버전 Matrix 검사 도구는 Akeroyd et al.(2015)의 권고 사항을 따라 한국어 문장 목록을 구성 및 최적화하였고(Kim & Lee, 2018; Yi, 2016), 20~30대의 건청 성인을 대상으로 음장 matrix 문장인지역치를 반복 측정하여 검사-재검사 신뢰도를 확인하였다(Jung et al., 2021). 신뢰도 분석 결과 건청 성인을 대상으로 반복된 측정 결과 간 일치 정도가 적절하였다. 그러나 20~30대의 건청 성인에 비해 노인 대상자의 소음 하 어음인지 능력에 개인 차이가 큰 편이므로(Humes, 2021; Nuesse et al., 2018) 노인을 대상으로 검사-재검사 신뢰도 및 타당도가 적절한지 확인하는 것이 필요하다. 특히 건청 성인보다 건청 노인이 포락선-변조 소음(envelope-modulated noise)과 같이 시간에 따라 강도가 변동하는 소음(fluctuation noise) 하에서 어음인지 시 순간적으로 빠르게 glimpsing하는 것이 어려워(Bologna et al., 2018; Meister et al., 2013) 검사 시 비변동 소음인 어음스펙트럼 소음뿐 아니라 변동 소음도 함께 제시하는 것을 고려해볼 수 있다.

따라서 본 연구에서는 건청 노인과 난청 노인을 대상으로 어음스펙트럼 소음과 변동 소음 상황에서 한국어 Matrix 문장인지 검사를 반복 시행하여 검사-재검사 신뢰도를 확인하고, 한국어 HINT 문장인지검사 결과와 비교하여 검사 도구의 공인 타당도를 검증하고자 하였다. 검사-재검사 신뢰도 확인 시 급내상관계수(intraclass correlation coefficient, ICC)를 구하여 상대적 신뢰도를 확인하였고, 측정오차의 발생을 최대한 배제하기 위해 절대적 신뢰도(측정오차 신뢰도)인 표준측정오차(standard error of measurement, SEM)와 최소감지변화값(minimal detectable change, MDC)을 함께 확인하였다.

MATERIALS AND METHODS

연구 대상

본 연구에는 60세 이상의 40명의 노인 대상자가 참여하였다. 대상자에게 한국판 몬트리올 인지평가(Korean version of Montreal Cognitive Assessment, MoCA-K)를 시행한 결과 인지 기능 저하를 보이는 대상자가 없음을 확인하였다(Lee et al., 2008). 대상자 모두 이과 및 신경학적 병력이 없었으며, 청력 검사기(AudioStar pro; Grason-Stadler, Eden Prairie, MN, USA), 헤드폰(TDH-39P; Telephonics, Farmingdale, NY, USA)을 사용하여 순음청력검사를 시행하였다. 세계보건기구(World Health Organization, WHO)에서 제안한 청력손실등급(WHO-proposed hearing-impairment grade)에 기준하였을 때(Humes, 2019; Stevens et al., 2013) 20명은 양이 모두 정상청력을 보여(0.5, 1, 2, 4 kHz 평균 순음청력역치가 20 dB hearing level [HL] 미만) 건청 노인군(elderly group with normal hearing, ENH)이라 명명하였다. 나머지 20명은 위의 WHO 청력손실등급을 기준으로 하였을 때 양이 청력이 경도부터 심도 난청(mild to severe)을 보여 난청 노인군(elderly group with hearing impairment, EHI)이라 명명하였다.

Table 1에는 ENH의 성별, 나이, MoCA-K 점수, 양쪽 귀의 순음역치평균(0.5, 1, 2, 4 kHz 평균) 정보를, Table 2에는 EHI의 정보를 나타내었다. ENH의 오른쪽 귀의 경우 0.25~8 kHz 옥타브 단위 주파수 순서대로 평균 16.75, 13.00, 14.00, 12.25, 19.00, 41.75 dB HL의 순음청력역치를 보였고, 왼쪽 귀의 경우 동일한 주파수 순서대로 평균 16.75, 12.75, 12.50, 12.50, 18.50, 35.25 dB HL의 순음청력역치를 보였다. EHI의 청력검사 결과, 오른쪽 귀의 경우 위와 동일한 주파수 순서대로 평균 26.00, 26.00, 31.50, 38.50, 53.75, 70.25 dB HL의 순음청력역치를 보였고, 왼쪽 귀의 경우 동일한 순서대로 평균 29.75, 28.00, 34.25, 44.25, 60.75, 72.00 dB HL의 순음청력역치를 보였다. 독립표본 *t* 검정 결과 ENH와 EHI 그룹 간 나이는 유의하게 다르지 않았으며($t[31.65] = 0.09$), 대상자 모두 연구에 참여하기 전 본 연구의 목적 및 절차에 대한 설명을 듣고 연구 참여에 동의하였다.

연구 절차

본 연구에서는 변동형 검사 절차를 통해 한국어 Matrix 문장인지검사의 검사-재검사 신뢰도를 측정하고, 한국어 HINT 검사 결과와 비교하여 타당도를 확인하고자 하였다. 모든 실험은 국제표준에서 권고하는 소음허용수준(International Organization for Standardization, 2012)을 만족하는 방음실에서 실시하였다.

Table 1. Information of ENH subjects

Subject	Gender	Age (yr)	MoCA-K	Rt PTA (dB HL)	Lt PTA (dB HL)
ENH1	F	60	27	20.00	20.00
ENH2	F	61	29	13.75	13.75
ENH3	F	64	29	12.50	10.00
ENH4	F	61	27	15.00	20.00
ENH5	F	66	26	16.25	10.00
ENH6	F	67	25	20.00	20.00
ENH7	F	66	28	12.50	20.00
ENH8	F	67	27	12.50	7.50
ENH9	F	61	27	13.75	10.00
ENH10	F	63	28	13.75	10.00
ENH11	F	65	27	8.75	20.00
ENH12	F	69	27	7.50	7.50
ENH13	M	68	27	20.00	17.50
ENH14	F	71	27	13.75	16.25
ENH15	M	70	26	13.75	16.25
ENH16	F	72	27	20.00	20.00
ENH17	M	74	29	20.00	20.00
ENH18	F	64	30	13.75	8.75
ENH19	F	67	28	10.00	6.25
ENH20	M	68	28	13.75	7.50
Mean		66.20 ± 3.90	27.45 ± 1.19	14.56 ± 3.79	14.06 ± 5.39

Values are presented as mean ± standard deviation or number. ENH: elderly with normal hearing, MoCA-K: Korean version of Montreal Cognitive Assessment, Rt: right, PTA: puretone threshold average across 0.5, 1, 2, 4 kHz, HL: hearing level, Lt: left, F: female, M: male

한국어 Matrix 문장인지검사

한국어 Matrix 문장인지검사를 시행하기 위해 Oldenburg Measurement Applications (OMA) software (HörTech GmbH, Oldenburg, Germany) (Brand & Kollmeier, 2002)를 노트북(gram: LG electronics, Seoul, Korea)에 설치한 후, 대상자의 귀로부터 1미터 거리, 정면 방향에 놓인 라우드스피커(SC-M53; DENON, Japan)와 Fireface UCX 디지털 아날로그 변환기(RME, Haimhausen, Germany)를 통해 matrix 문장(목록당 문장 20개)을 제시하였다. OMA 소프트웨어에서는 대상자가 각 문장에 보인 반응에 따라 최대우도 추정(maximum likelihood estimator)에 근거하여(Brand & Kollmeier, 2002) 그 다음 문장의 강도가 변동 조절된다. Matrix 검사 절차대로 소음레벨은 65 dB SPL에 고정하여 제시하였고, 대상자의 반응에 따라 문장레벨을 조절하여 50% 소음 하 문장인지에 필요한 SNR을 측정하여 speech reception threshold (SRT) 결과를 도출하였다.

Table 2. Information of EHI subjects

Subject	Gender	Age (yr)	MoCA-K	Rt PTA (dB HL)	Lt PTA (dB HL)
EHI1	M	74	23	31.25	30.00
EHI2	M	64	28	30.00	31.25
EHI3	F	77	25	26.25	31.25
EHI4	M	75	23	32.50	28.75
EHI5	M	64	24	33.75	33.75
EHI6	M	73	25	26.25	31.25
EHI7	M	60	23	32.50	38.75
EHI8	M	75	23	51.25	46.25
EHI9	F	78	23	57.50	48.75
EHI10	M	70	24	30.00	47.50
EHI11	F	72	23	42.50	38.75
EHI12	M	77	23	31.25	30.00
EHI13	M	73	26	40.00	38.75
EHI14	M	62	27	33.75	38.75
EHI15	F	66	25	52.50	47.50
EHI16	F	68	25	33.75	78.75
EHI17	M	62	28	50.00	55.00
EHI18	F	63	29	26.25	57.50
EHI19	M	61	27	43.75	42.50
EHI20	M	72	27	43.75	41.25
Mean		69.30 ± 6.04	25.05 ± 2.01	37.44 ± 9.58	41.81 ± 12.15

Values are presented as mean ± standard deviation or number. EHI: elderly with hearing impairment, MoCA-K: Korean version of Montreal Cognitive Assessment, Rt: right, PTA: puretone threshold average across 0.5, 1, 2, 4 kHz, HL: hearing level, Lt: left, M: male, F: female

소음 하 문장인지 측정을 위해 세 가지 종류의 소음을 배경 소음으로 제시하였다. 첫 번째 소음은 matrix 문장의 스펙트럼을 가진 어음스펙트럼 소음(speech-shaped noise, SSN)이며, 두 번째 소음은 어음의 스펙트럼과 변동적 특성을 갖도록 제작된 International Collegium of Rehabilitative Audiology (ICRA) 소음이었으며, 세 번째 소음은 International Speech Test Signal (ISTS)이었다. Matrix 검사 도구에 포함되어 있는 SSN 소음은 총 300개의 matrix 문장 음원을 중첩(superimpose)하여 만든 소음으로, SSN 소음과 matrix 문장 음원의 장기 평균 어음스펙트럼이 유사하다는 특성을 가진다. ICRA 소음(Dreschler et al., 2001)의 경우 여성 화자가 녹음한 영어 어음을 필터링한 표준 어음 스펙트럼을 가지도록 제작한 소음이다. SSN 소음이 어음의 장기평균어음스펙트럼을 가지고 있으나 강도는 일정한 것에 비해 ICRA 소음은 어음의 장기평균어음스펙트럼을 가지고 있으며, 시간에 따라 강도가 변화하는 어음의 포락선(envelope) 형태를 가지는 광대역 소음이다. 선

행 연구(Hochmuth et al., 2015)에 따르면 SSN 소음에 비해 ICRA 소음 제시 시 강도 변화 시 목표 어음에 대해 glimpsing (Jensen & Bernstein, 2019)이 가능하여 약 10 dB 가량의 신호대잡음비 혜택이 있었다. ISTS 소음은 International Standards for Measuring Advanced Digital Hearing Aids Working group에 의해 제작된 소음으로 보청기의 어음처리 과정 분석을 위해 개발되었다(Holube et al., 2010). ISTS 소음은 언어에 상관없이 적용 가능하게 하기 위해 6개의 서로 다른 언어를 모국어로 사용하는 여성 화자가 모국어로 동일한 이야기("The north wind and the sun")를 읽도록 녹음하고 100~600 msec로 분할한 후 무작위 순서로 연결된 소음이다.

선행 연구에서 Matrix 문장인지검사 시 최소 2개의 연습 과정을 거친 후 소음 하 문장인지 결과가 안정적이었다고 보고한 바 있어(Wagener & Brand, 2005) 본 연구에서도 SSN 소음 조건에서 2개의 문장 목록을 통해 연습 과정(practice)을 가진 후 실험을 진행하였다. 순서에 따른 학습 효과를 고려하여 라틴스퀘어 실험 설계에 따라 세 가지 소음(SSN, ICRA, ISTS)을 제시하였고, 검사와 재검사는 2주 간격을 두고 시행하였다.

한국어 HINT 검사

한국어 HINT 검사를 위해 상용화된 Bio-logic 사의 HINT Pro 장비(Bio-logic® sys., Mundelein, IL, USA)를 사용하였다. HINT 검사 시 목표 문장은 정면 방향 1미터 거리에 위치한 스피커를 통해 제시하였고, 65 dBA 고정된 강도에서 어음스펙트럼 소음을 제시하였다. 한국어 Matrix 검사와 마찬가지로 한국어 HINT 검사 역시 변동형 절차에 따라 검사 시 대상자의 정반응 혹은 오반응에 따라 목표 문장의 강도가 조절되었다. 일반적인 HINT 검사 절차(Nilsson et al., 1994)에 따라 소음 환경은 정면(0°), 우측 소음(90°), 좌측 소음(-90°) 조건으로 나누어 검사하였고, 각 소음 조건(정면, 좌측, 우측 소음)에서 50% 소음 하 문장인지에 필요한 SNR을 확인하여 소음 하 인지 능력을 평가하였다. 정면, 좌측, 우측에서 소음이 제시되었을 때 구한 SNR을 통해 복합신호대잡음(composite SNR)을 구하였다.

$$\text{Composite SNR} = [(\text{정면 소음 SNR} \times 2) + \text{우측 소음 SNR} + \text{좌측 소음 SNR}] / 4 \quad [1]$$

통계 분석

수집된 자료는 Statistical Product and Service Solution (SPSS version 25.0; IBM Corp., Armonk, NY, USA)을 이용하여 분석하였다. 상대적 신뢰도 검증을 위해 이원배치변량(임의)효과모형(two-way random effects model)을 기준으로 ICC(2,1) 결과를 확인하여(ICC 모델 2, 평가자 1명 측정) 반복 측정된 결과의 일치성(agreement)을 확인하였다. ICC는 0부터 1까지의 수

로 표기하며, 1에 가까울수록 일치도의 정도가 높음을 의미한다(Shrout & Fleiss, 1979). ICC(2,1) 상관계수가 0.50~0.75이면 검사-재검사 신뢰도가 보통(moderate), ICC 계수가 0.75~0.90이면 검사-재검사 신뢰도가 우수(good), ICC 계수가 0.90 초과이면 검사-재검사 신뢰도가 매우 우수(excellent)하다고 본다(Koo & Li, 2016). 상대적 신뢰도 검증 외에 절대적 신뢰도 검증을 위해 반복된 평가 시 우연히 발생 가능한 측정오차를 확인하고자 측정의 표준오차인 SEM과 MDC를 구하였다. SEM은 검사-재검사 결과의 표준편차(standard deviation, SD)와 ICC의 오차 변량의 제곱근을 사용하여 구하였다.

$$\text{SEM} = \text{SD} \times \sqrt{[1 - \text{ICC coefficient}]} \quad [2]$$

MDC는 95% 신뢰구간(confidence interval, CI)에서 $\text{MDC} = 1.96 \times \sqrt{2} \times \text{SEM}$ 을 이용하여 산출하였다(Haley & Fragala-Pinkham, 2006). Bland-Altman 도표(Bland & Altman, 1986; Giavarina, 2015)를 통해 검사-재검사의 두 검사 결과의 일치 정도를 시각적으로 확인하였으며, 95% 일치한계 수준(limits of agreement, LoA)을 통해 검사-재검사 측정 결과 간 차이의 95%가 놓이는 범위에 대한 정보를 확인하였다. 공인 타당도(concurrent validity) 검증을 위해 피어슨 상관 분석을 통해 한국어 Matrix 소음 하 문장인지검사 결과와 HINT 결과 간 상관성을 확인하였다.

RESULTS

검사-재검사 신뢰도 검증 결과

Table 3은 ENH와 EHI 그룹을 대상으로 측정한 검사-재검사 결과를 보여준다. ENH 그룹을 대상으로 비변동성 어음스펙트럼 소음인 SSN, 어음의 변동성 특징을 가진 ICRA 소음, ISTS 소음을 통해 측정한 검사-재검사 평균 SRT는 다음과 같다. ENH에게 SSN 소음을 제시하여 측정한 첫 검사의 평균 SRT는 -6.24 dB SNR (SD: 1.10, range: -8.30 to -4.30 dB SNR)이었고, 2주 후 측정한 재검사에서의 평균 SRT는 -6.63 dB SNR (SD: 1.08, range: -8.90 to -4.70 dB SNR)이었다. ENH에게 ICRA 소음을 제시하고 측정한 검사-재검사 평균 SRT는 -15.21 dB SNR (SD: 3.42, range: -20.30 to -7.60 dB SNR)과 -15.89 dB SNR (SD: 3.22, range: -20.90 to -10.30 dB SNR)이었다. ISTS 소음을 통해 측정한 ENH의 검사-재검사 평균 SRT는 -15.41 dB SNR (SD: 3.17, range: -20.30 to -9.00 dB SNR)과 -16.40 dB SNR (SD: 3.05, range: -23.70 to -12.20 dB SNR)이었다.

EHI 그룹을 대상으로 SSN을 제시하여 측정한 결과를 살펴보면 첫 번째 검사에서 측정한 평균 SRT는 -3.26 dB SNR (SD: 2.80, range: -6.60 to 3.50 dB SNR), 2주 후 재검사에서 측정

Table 3. Reliability indices of the Korean matrix sentence-in-noise test of ENH and EHI groups

Noise type	First test (dB SNR)	Second test (dB SNR)	ICC (95% CI)	SEM	MDC
ENH					
SSN	-6.24 ± 1.10	-6.63 ± 1.08	0.84 (0.66~0.94)	0.44	1.21
ICRA	-15.21 ± 3.42	-15.89 ± 3.22	0.94 (0.83~0.98)	0.81	2.24
ISTS	-15.41 ± 3.17	-16.40 ± 3.05	0.89 (0.67~0.96)	1.03	2.86
EHI					
SSN	-3.26 ± 2.80	-3.24 ± 2.50	0.92 (0.79~0.97)	0.74	2.06
ICRA	-7.11 ± 5.32	-7.26 ± 4.98	0.98 (0.94~0.99)	0.72	1.99
ISTS	-5.67 ± 7.02	-6.81 ± 6.50	0.95 (0.86~0.98)	1.49	4.15

Values are presented as mean SRT ± standard deviation unless otherwise indicated. ENH: elderly with normal hearing, EHI: elderly with hearing impairment, SNR: signal-to-noise ratio, ICC: intraclass correlation coefficient, CI: confidence interval, SEM: standard error of measurement, MDC: minimal detectable change, SSN: speech-shaped noise, ICRA: International Collegium of Rehabilitative Audiology, ISTS: International Speech Test Signal, SRT: speech reception threshold

한 평균 SRT는 -3.24 dB SNR (SD: 2.50, range: -7.30 to 1.70 dB SNR)이었다. ICRA 소음을 통해 측정된 EHI의 검사-재검사 평균 SRT는 -7.11 dB SNR (SD: 5.32, range: -16.40 to 2.80 dB SNR)과 -7.26 dB SNR (SD: 4.98, range: -15.90 to 3.20 dB SNR)이었다. 마지막으로 ISTS 소음을 통해 측정된 EHI의 검사-재검사 평균 SRT는 -5.67 dB SNR (SD: 7.02, range: -17.00 to 6.70 dB SNR)과 -6.81 dB SNR (SD: 6.50, range: -15.80 to 5.40 dB SNR)이었다.

상대적 신뢰도 검증을 위해 급내상관계수 ICC(2,1)를 확인하여 검사-재검사 결과 간 일치성 정도를 확인하였다. 본 연구에서 ENH 그룹에게 SSN을 제시하였을 때 검사-재검사 신뢰도 ICC (95% CI)는 0.84 (0.66~0.94), SEM은 0.44, MDC는 1.21이었으며, ICRA 소음 제시 시 ENH 그룹의 ICC (95% CI)는 0.94 (0.83~0.98), SEM은 0.81, MDC는 2.24였다. ISTS 소음을 통해 측정된 ENH 그룹의 ICC (95% CI)는 0.89 (0.67~0.96), SEM은 1.03, MDC는 2.86이었다. SSN 소음 제시 시 EHI 그룹의 경우 검사-재검사 신뢰도 ICC (95% CI)는 0.92 (0.79~0.97), SEM은 0.74, MDC는 2.06이었다. ICRA 제시 시 EHI 그룹의 검사-재검사 신뢰도 ICC (95% CI)는 0.98 (0.94~0.99), SEM은 0.72, MDC는 1.99였다. ISTS 제시 시 EHI 그룹의 검사-재검사 신뢰도 ICC (95% CI)는 0.95 (0.86~0.98), SEM은 1.49, MDC는 4.15였다. ICC(2,1) 상관계수가 0.75~0.90이면 검사-재검사 신뢰도가 우수(good), ICC 계수가 0.90 초과이면 검사-재검사 신뢰도가 매우 우수(excellent)하다고 판독하므로(Koo & Li, 2016), 본 연구 결과에서 두 그룹의 검사-재검사 신뢰도가 소음의 종류에 상관없이 적절하였음을 확인하였다. MDC는 두 결과 간 점수의 변화가 일어났을 때 단순히 오류 때문에 생긴 변화가 아닌 대상자 개인에게서 생긴 변화라고 간주할 수 있는 최소한의 점수변화량을 의미한다. 본 연구에서 확인한 두 그룹의 MDC

수치 상 전반적으로 무작위측정오류(random measurement error)의 정도가 수용할만한 수준이나, ENH 그룹에 비해 EHI 그룹의 경우 ISTS 변동 소음 제시 시 MDC, 즉 대상자 개인에게서 생긴 변화라고 간주할 수 있는 최소한의 점수변화량 정도가 더 큼을 확인하였다.

Figure 1은 ENH 그룹에게 SSN, ICRA, ISSTS 소음을 제시하여 검사-재검사 결과의 일치 정도를 확인하기 위하여 구한 Bland & Altman plot이며, Figure 2는 EHI 그룹에게 동일한 소음을 제시하여 구한 Bland & Altman plot이다. ENH 그룹을 대상으로 구한 검사-재검사 일치한계(LoA)의 범위는 SSN 제시 시 -1.94~2.73, ICRA 소음의 경우 -2.54~3.90, ISTS의 경우 LoA 범위가 -2.49~4.47이었다. EHI 그룹을 대상으로 구한 LoA 범위는 SSN 제시 경우 -3.01~3.05, ICRA 소음의 경우 -4.89~4.60, ISTS의 경우 -10.53~8.26이었다.

공인 타당도 검증 결과

피어슨 상관 분석을 통해 SSN, ICRA, ISTS 소음을 통해 측정된 SRT 간 유의한 상관성이 있는지 확인하였다. 분석 결과 SSN을 통해 구한 Matrix 검사에서 SRT가 높을수록 정면 혹은 좌우 방향에서 측정된 SRT가 유의하게 높았다($r = 0.58\sim0.72, p < 0.001$). ICRA 소음 혹은 ISTS 소음과 같은 변동 소음을 제시하여 측정된 Matrix 검사에서 SRT가 높을수록 정면 혹은 좌우 방향에서 측정된 SRT 또한 유의하게 높았다($r = 0.52\sim0.68, p < 0.001$). 그 외에 Matrix 소음 하 검사에서 세 가지 소음을 통해 측정된 결과 간 상관성을 분석한 결과, SSN을 통해 구한 SRT가 높을수록 ICRA 소음을 제시하고 측정된 SRT ($r = 0.87, p < 0.001$) 혹은 ISTS 소음을 제시하고 측정된 SRT ($r = 0.72, p < 0.001$) 결과가 유의하게 높았다. ICRA 소음 하에서 구한 SRT가 높을수록 ISTS 소음 하에서 구한 SRT ($r = 0.87, p <$

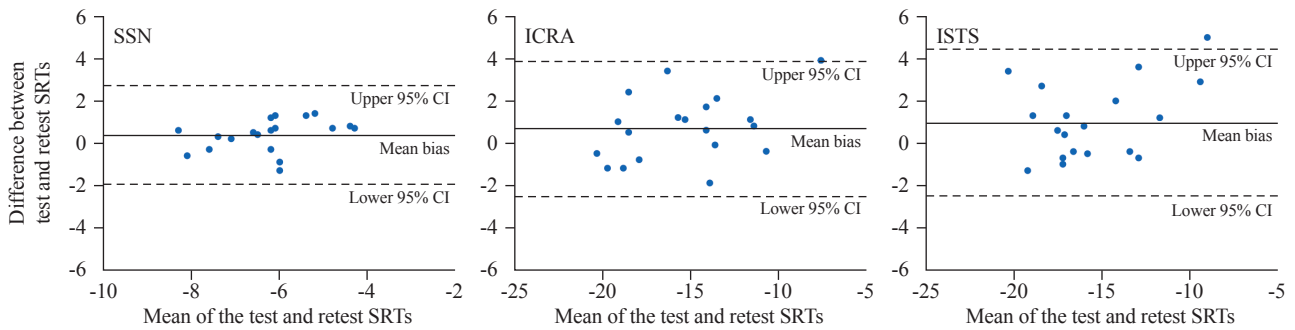


Figure 1. Bland-Altman plots for elderly with normal hearing (ENH) group showing the difference between the test and retest measurements against the mean of the two measurements. SRT: speech reception threshold, SSN: speech-shaped noise, CI: confidence interval, ICRA: International Collegium of Rehabilitative Audiology, ISTS: International Speech Test Signal.

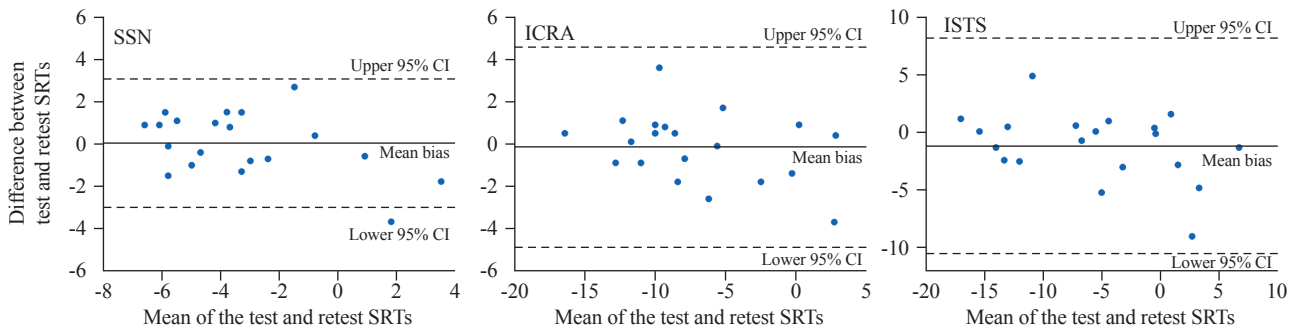


Figure 2. Bland-Altman plots for elderly with hearing impairment (EHI) group displaying the difference between the test and retest measurements against the mean of the two measurements. SRT: speech reception threshold, SSN: speech-shaped noise, CI: confidence interval, ICRA: International Collegium of Rehabilitative Audiology, ISTS: International Speech Test Signal.

0.001) 결과가 유의하게 높았다.

DISCUSSIONS

소음 하 듣기 상황에서 청자의 어려움을 측정하기 위해서는 검사 도구의 임상 목적에 맞게 검사 결과의 신뢰도 및 타당도 등을 검증하는 것이 필요하다. 한국어 Matrix 소음 하 문장인지검사 도구의 경우 헤드폰 혹은 음장스피커를 사용하는 검사 조건에서 20~30대의 건청 성인을 대상으로 검사 도구의 신뢰도 및 타당도 등을 검증한 바 있으나(Jung et al., 2021; Kim & Lee, 2018) 노인 혹은 난청인을 대상으로 신뢰도와 타당도를 검증한 연구가 제한적이었다. 노인 대상자의 소음 하 어음인지 능력을 반복 평가할 일이 빈번하므로 반복하여 측정한 평가 결과가 다를 때 우연히 발생한 측정오차에 의한 것인지, 실제로 결과가 변화한 것인지 검증하는 것이 중요하다. 따라서 본 연구에서는 건청 노인인군과 난청 노인을 대상으로 한국어 Matrix 검사의 검사-재검사 일치성 및 신뢰도, 공인 타당도가 적절한지 확인하였다.

먼저, 한국어 Matrix 검사-재검사 결과의 일치성(agreement)을 확인하기 위해 상대적 신뢰도 지수인 급내상관계수 ICC를 확

인하였다. ICC가 클수록 높은 신뢰도를 의미하며 1은 완벽한 일치도를 0은 일치도가 전혀 없는 것을 의미하며, 보통 ICC(2,1) 상관계수가 0.50~0.75이면 검사-재검사 신뢰도가 보통(moderate), ICC 계수가 0.75~0.90이면 신뢰도가 우수(good), ICC 계수가 0.90 초과이면 신뢰도가 매우 우수(excellent)하다고 본다(Koo & Li, 2016). 선행 연구(Jung et al., 2021)에서 20~30대 건청 성인을 대상으로 동일한 방법으로 한국어 Matrix 검사-재검사 신뢰도 결과의 일치성을 확인한 결과 ICC (95% CI) 계수가 0.84 이상으로 우수한 신뢰도를 보였다. 본 연구 데이터 분석 결과 세 가지 소음 제시 시 ENH와 EHI 그룹의 ICC 계수 범위가 0.84~0.98 범위이었으므로 노인군의 검사-재검사 결과의 신뢰도가 우수함을 확인하였다. 따라서 건청 성인뿐 아니라 건청 노인, 난청 노인을 대상으로 한국어 Matrix 검사를 시행할 때 적절한 검사-재검사 일치율을 기대해볼 수 있었다.

본 연구에서는 한국어 Matrix 검사 도구의 신뢰도 측정을 위해 상대적 신뢰도뿐 아니라 절대적 신뢰도 지수인 SEM과 MDC 값을 구하였다. 여기서 MDC 값은 두 결과 간 점수의 변화가 일어났을 때 반복된 평가 시 우연히 발생 가능한 측정오차 때문에 생긴 변화가 아니라 대상자 개인에게서 생긴 변화라고 간주

할 수 있는 최소한의 점수 변화량을 의미한다. 건청 성인을 대상으로 한국어 Matrix 검사의 절대적 신뢰도 지수를 구한 결과 (Jung et al., 2021) SSN 소음에 비해(SEM: 0.33, MDC: 0.92) ICRA, ISTS와 같은 변동 소음 제시 시 SEM, MDC 값이 비교적 더 컸으며(SEM: 약 1.1, MDC: 약 2.9), 소음 종류에 상관없이 전반적으로 무작위 측정오류의 정도가 수용할만하다고 보고하였다. 건청 성인을 대상으로 구한 SEM, MDC 값(Jung et al., 2021)과 본 연구에서 측정된 ENH와 EHI 그룹의 SEM, MDC 값과 비교하였을 때(Table 3), 건청 성인에 비해 건청 노인과 난청 노인 모두 SEM과 MDC 값이 비교적 컸음을 확인할 수 있었다. 세 가지 소음 중 특히 ISTS 소음 제시 시 무작위 측정오류 정도가 컸다(SEM: 1.49, MDC: 4.15). Meister et al.(2013) 역시 비변동 소음, ICRA 소음, ISTS 소음을 통해 Matrix 소음하 문장인지검사를 진행하였다. 분석 결과 소음 종류에 상관없이 성인에 비해 노인의 인지도가 유의하게 낮았으며, 특히 그룹 간 차이가 변동 소음 하에서 어음인지 시 더 두드러졌다고 밝혔다. 본 연구에서 EHI 그룹의 경우 ICRA 소음에 비해 ISTS 소음 제시 시 개인 간 차이뿐 아니라 개인 내 변화가 컸음을 알 수 있다. 이와 같은 결과를 보인 이유는 ICRA 소음의 경우 어음의 장기평균 어음스펙트럼과 어음의 포락선 정보는 소음 내에 포함되어 있지만 경쟁 어음으로서의 의미적 차폐 효과를 주거나 경쟁 어음 내 시간미세구조(temporal-fine-structure)는 가지고 있지 않은 특징 때문일 수 있다(Summers & Roberts, 2020). ISTS 소음의 경우 6개의 서로 다른 언어를 모국어로 사용하는 여성 화자의 이야기기가 무작위 순서로 연결된 소음으로 경쟁 어음으로서의 의미적 차폐를 보이지는 않으나 경쟁 화자의 어음이 가지는 시간미세구조는 가지고 있기 때문에 이러한 특징이 노인 대상자에게 영향을 주었다고 볼 수 있다. Füllgrabe (2013), Füllgrabe et al.(2018)은 임상에서의 청력역치가 정상이어도 청자의 나이에 따라 시간미세구조 민감도가 떨어지며 이것은 소음 하 어음인지 능력 저하와 관련성이 있음을 보고하였다.

본 연구의 결과를 통해 한국어 Matrix 검사의 검사-재검사 일치율 및 신뢰도가 우수하였고, HINT 결과와의 상관성을 고려하였을 때 공인 타당도 또한 적절하다고 볼 수 있다. HINT 검사의 경우 총 목표 문장이 250개이지만 Matrix 검사 도구의 경우 생성 가능한 문장의 수가 10,000개이며, 문장의 문맥적 단서가 비교적 낮고 50% SNR을 자동으로 찾아주는 검사의 편리함이 있다. 방향성 마이크로폰 보청기 착용인을 대상으로 영어 Matrix 검사와 HINT 검사 간 심리음향곡선을 비교 분석한 결과(Harianawala et al., 2019), 영어 Matrix 검사와 HINT 검사 모두 방향성 마이크로폰 사용 혜택 정도를 측정하는 민감도가 유사하였으므로 두 검사 모두 임상적 의의가 좋다고 밝힌 바 있다. HINT 검사의 경우 어음스펙트럼 소음만을 사용해

야 하는 것에 비해 Matrix 검사의 경우 다양한 소음 종류를 이용할 수 있다. Matrix 검사에서 변동 소음과 비변동 소음을 모두 활용할 경우, 비변동 소음에 비해 변동 소음 제시 시 일시적으로 낮은 진폭을 나타내는 순간 신호대소음비 상승하여 목표 어음을 더 많이 인지할 수 있게 되는 glimpsing, 즉 변동소음혜택(fluctuating masker benefit, FMB)을 정량화할 수 있다. Jung et al.(2021)의 경우 건청 성인의 FMB가 약 12 dB 가량이라고 보고하였고, 본 연구에서 측정된 건청 성인의 FMB는 약 9~10 dB, 난청 노인의 FMB는 약 3~4 dB 가량이었다. Matrix 검사 시 임상 진단 혹은 연구 목적에 따라 소음의 종류를 선택하여 검사에 활용한다면 보다 편하게 FMB를 측정할 수 있겠다.

본 연구는 여러 가지 제한점을 가지고 있다. 첫 번째로는 그룹당 대상자 수가 20명으로 제한적이었다는 점이다. 선행 연구(Shayanmehr et al., 2015)에서 건청 성인을 대상으로 Persian Quick Speech in Noise Test를 2주 간격으로 시행하여 검사 도구의 검사-재검사 신뢰도와 타당도를 밝혔다. 노인 대상자의 경우 개인 차이가 많음을 고려하였을 때보다 많은 대상자를 통해 검증을 추가하는 것이 필요하겠다. 두 번째로는 검사-재검사 기간을 2주 간격으로 하였다는 점이다. 추후 1개월, 3개월 등의 장기적 간격의 검사-재검사 결과를 포함하여 검사-재검사 신뢰도를 확인하는 것이 필요하겠다.

본 연구에서 건청 및 난청 노인을 대상으로 한국어 Matrix 소음 하 문장인지검사의 검사-재검사 일치율, 신뢰도 및 타당도가 적절한지 확인하고자 하였다. 건청 노인과 난청 노인 그룹 모두 상대적 신뢰도 및 일치율이 우수한 편이었으며, 건청 노인에 비해 난청 노인의 경우 세 가지 소음 중 ISTS 소음 제시 시 우연 변동에 의해 발생하는 측정오차가 비교적 큰 편이었다. 성인 외에 노인을 대상으로 신뢰도와 타당도를 검증한 검사 도구가 부족한 편이므로 향후 노인의 소음 하 인지평가 및 재활 계획 등의 목적으로 본 연구 결과를 유용하게 활용할 수 있겠다.

중심 단어: 신뢰도, 타당도, 한국어 Matrix 문장인지검사, 소음 하 어음인지.

Ethical Statement

This study was approved by the Institutional Review Board of Hallym University of Graduate Studies (HUGSAUD #256198).

Acknowledgments

N/A

Declaration of Conflicting Interests

There is no conflict of interests.

Funding

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5A2A01051014).

Author Contributions

Conceptualization: Yukyung Jung, Jae Hee Lee. Data collection and Formal analysis: Yukyung Jung. Funding acquisition: Jae Hee Lee. Writing—original draft: Yukyung Jung. Writing—review & editing: all authors. Approval of final manuscript: Jae Hee Lee.

ORCID iD

Yukyung Jung <https://orcid.org/0000-0002-9774-9899>
 Jaeho Han <https://orcid.org/0000-0001-8364-0780>
 Hyo Jung Choi <https://orcid.org/0000-0002-8193-4010>
 Jae Hee Lee <https://orcid.org/0000-0002-4152-6434>

REFERENCES

- Akeroyd, M. A., Arlinger, S., Bentler, R. A., Boothroyd, A., Dillier, N., Dreschler, W. A., et al. (2015). International Collegium of Rehabilitative Audiology (ICRA) recommendations for the construction of multilingual speech tests. ICRA working group on multilingual speech tests. *International Journal of Audiology, 54* Suppl 2, 17-22.
- Anderson, S. & Kraus, N. (2010). Sensory-cognitive interaction in the neural encoding of speech in noise: A review. *Journal of the American Academy of Audiology, 21*(9), 575-585.
- Bland, J., M. & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 327*(8476), 307-310.
- Bologna, W. J., Vaden, K. I., Jr, Ahlstrom, J. B., & Dubno, J. R. (2018). Age effects on perceptual organization of speech: Contributions of glimpsing, phonemic restoration, and speech segregation. *The Journal of the Acoustical Society of America, 144*(1), 267-281.
- Brand, T. & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America, 111*(6), 2801-2810.
- Chang, H. P., Ho, C. Y., & Chou, P. (2009). The factors associated with a self-perceived hearing handicap in elderly people with hearing impairment—results from a community-based study. *Ear and Hearing, 30*(5), 576-583.
- Ciorba, A., Bianchini, C., Pelucchi, S., & Pastore, A. (2012). The impact of hearing loss on the quality of life of elderly adults. *Clinical Interventions in Aging, 7*, 159-163.
- Dayasiri, M., Dayasena, R., Jayasuriya, C., Perera, D., Kuruppu, K. A., & Peris, M. (2011). Quantitative analysis of the effect of the demographic factors on presbycusis. *The Australasian Medical Journal, 4*(3), 118-122.
- Dreschler, W. A., Verschuure, H., Ludvigsen, C., & Westermann, S. (2001). ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. International collegium for rehabilitative audiology. *Audiology, 40*(3), 148-157.
- Füllgrabe, C. (2013). Age-dependent changes in temporal-fine-structure processing in the absence of peripheral hearing loss. *American Journal of Audiology, 22*(2), 313-315.
- Füllgrabe, C., Şek, A. P., & Moore, B. (2018). Senescent changes in sensitivity to binaural temporal fine structure. *Trends in Hearing, 22*, 1-16.
- Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica, 25*(2), 141-151.
- Haley, S. M. & Fragala-Pinkham, M. A. (2006). Interpreting change scores of tests and measures used in physical therapy. *Physical Therapy, 86*(5), 735-743.
- Harianawala, J., Galster, J., & Hornsby, B. (2019). Psychometric comparison of the Hearing in Noise Test and the American English Matrix Test. *Journal of the American Academy of Audiology, 30*(4), 315-326.
- Hochmuth, S., Kollmeier, B., Brand, T., & Jürgens, T. (2015). Influence of noise type on speech reception thresholds across four languages measured with matrix sentence tests. *International Journal of Audiology, 54* Suppl 2, 62-70.
- Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an International Speech Test Signal (ISTS). *International Journal of Audiology, 49*(12), 891-903.
- Humes, L. E. (2019). The World Health Organization's hearing-impairment grading system: An evaluation for unaided communication in age-related hearing loss. *International Journal of Audiology, 58*(1), 12-20.
- Humes, L. E. (2021). Factors underlying individual differences in speech-recognition threshold (SRT) in noise among older adults. *Frontiers in Aging Neuroscience, 13*, 702739.
- International Organization for Standardization. (2012). *Acoustics—Audiometric Test Methods—Part 3: Speech Audiometry (ISO 8253-3 2012)*. Geneva: International Organization for Standardization.
- Jensen, K. K. & Bernstein, J. (2019). The fluctuating masker benefit for normal-hearing and hearing-impaired listeners with equal audibility at a fixed signal-to-noise ratio. *The Journal of the Acoustical Society of America, 145*(4), 2113-2125.
- Jung, Y., Han, J. H., Choi, S., & Lee, J. H. (2021). Test-retest reliability of the Korean matrix sentence-in-noise recognition in sound-field testing condition. *Audiology and Speech Research, 17*(4), 344-351.
- Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 116*(4 Pt 1), 2395-2405.
- Kim, G. M. & Lee, J. H. (2020). Prediction of software-based sentence-in-noise recognition thresholds of hearing-impaired listeners. *Audiology and Speech Research, 16*(2), 140-146.
- Kim, K., Lee, J. H., & Bahng, J. (2021). Factors affecting the auditory working memory of the elderly. *Audiology and Speech Research, 17*(1), 35-43.
- Kim, K. H. & Lee, J. H. (2018). Evaluation of the Korean matrix sentence test: Verification of the list equivalence and the effect of word position. *Audiology and Speech Research, 14*(2), 100-107.
- Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Usilar, V., Brand, T., et al. (2015). The multilingual matrix test: Principles,

- applications, and comparison across languages: A review. *International Journal of Audiology*, 54 Suppl 2, 3-16.
- Koo, T. K. & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163.
- Lee, J. Y., Lee, D. W., Cho, S. J., Na, D. L., Jeon, H. J., Kim, S. K., et al. (2008). Brief screening for mild cognitive impairment in elderly outpatient clinic: Validation of the Korean version of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry and Neurology*, 21(2), 104-110.
- Lin, F. R., Niparko, J. K., & Ferrucci, L. (2011). Hearing loss prevalence in the United States. *Archives of Internal Medicine*, 171(20), 1851-1852.
- Meister, H., Schreitmüller, S., Grugel, L., Beutner, D., Walger, M., & Meister, I. (2013). Examining speech perception in noise and cognitive functions in the elderly. *American Journal of Audiology*, 22(2), 310-312.
- Moon, S. K., Kim, S. H., Mun, H. A., Jung, H. K., Lee, J. H., Choung, Y. H., et al. (2008). The Korean hearing in noise test. *International Journal of Audiology*, 47(6), 375-376.
- Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, 95(2), 1085-1099.
- Nuesse, T., Steenken, R., Neher, T., & Holube, I. (2018). Exploring the link between cognitive abilities and speech recognition in the elderly under different listening conditions. *Frontiers in Psychology*, 9, 678.
- Shayanmehr, S., Tahaei, A. A., Fatahi, J., Jalaie, S., & Modarresi, Y. (2015). Development, validity and reliability of Persian quick speech in noise test with steady noise. *Auditory and Vestibular Research*, 24(4), 234-244.
- Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428.
- Stevens, G., Flaxman, S., Brunskill, E., Mascarenhas, M., Mathers, C. D., Finucane, M., et al. (2013). Global and regional hearing impairment prevalence: An analysis of 42 studies in 29 countries. *European Journal of Public Health*, 23(1), 146-152.
- Summers, R. J. & Roberts, B. (2020). Informational masking of speech by acoustically similar intelligible and unintelligible interferers. *The Journal of the Acoustical Society of America*, 147(2), 1113-1125.
- Taylor, B. (2003). Speech-in-noise tests: How and why to include them in your basic test battery. *The Hearing Journal*, 56(1), 40-43.
- Vermiglio, A. J., Soli, S. D., Freed, D. J., & Fisher, L. M. (2012). The relationship between high-frequency pure-tone hearing loss, hearing in noise test (HINT) thresholds, and the articulation index. *Journal of the American Academy of Audiology*, 23(10), 779-788.
- Wagener, K. C. & Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. *International Journal of Audiology*, 44(3), 144-156.
- Wilson, R. H. (2011). Clinical experience with the words-in-noise test on 3430 veterans: Comparisons with pure-tone thresholds and word recognition in quiet. *Journal of the American Academy of Audiology*, 22(7), 405-423.
- Yi, D. (2016). Study on the construction and optimization of the Korean matrix sentence materials (Unpublished master's thesis). Seoul: Hallym University of Graduate Studies.