



# Test-retest Reliability of the Korean Matrix Sentence-in-noise Recognition in Sound-field Testing Condition

Yukyeong Jung<sup>1</sup>, Jae Ho Han<sup>1</sup>, Seokyu Choi<sup>1</sup>, Jae Hee Lee<sup>1,2</sup>

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

<sup>2</sup>HUGS Center for Hearing and Speech Research, Seoul, Korea

**Received:** July 9, 2021

**Revised:** August 31, 2021

**Accepted:** October 6, 2021

## 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:** leejaehlee@hallym.ac.kr

**Purpose:** This study aimed to measure the test-retest reliability of the Korean Matrix sentence-in-noise test in a sound-field testing condition. **Methods:** Fifty-five listeners with normal hearing participated. The sound-field speech-in-noise test was administered with a front loudspeaker by presenting noise at a fixed level (65 dB SPL) while adaptively adjusting the sentence level based on the listener's response. The speech reception thresholds (SRTs) corresponding to 50% sentence-in-noise intelligibility were obtained using speech-shaped noise (SSN), International Collegium of Rehabilitative Audiology (ICRA) noise, and International Speech Test Signal (ISTS) noise. For the test-retest reliability, each subject was tested twice in a 2-week interval. The test-retest reliabilities were judged by the intraclass correlation coefficient (ICC), the minimal detectable change (MDC) values, and Bland-Altman plots. **Results:** The test-retest reliabilities of the Korean Matrix sentence test were good (ICC, 0.84~0.86) across SSN, ICRA, ISTS noises. The MDC was 0.92 dB SNR for SSN and 2.83 and 2.98 dB SNR for ICRA and ISTS, indicating the random measurement error varied across the types of noise. **Conclusion:** The Korean Matrix sentence-in-noise test can be used as a reliable tool to assess speech-in-noise intelligibility in sound-field testing condition. Further study across various sessions would strengthen the present finding.

**Key Words:** Test-retest reliability, Korean Matrix sentence test, Evaluation of sound-field sentence recognition test in noise.

## INTRODUCTION

우리는 다양한 배경소음 속에서 일상생활을 하고 있으며, 일상적 배경소음은 청자의 어음인지 및 의사소통능력, 작업기억능력, 집중력 등 다양한 기능을 저하시킬 수 있다(Bronkhorst & Plomp, 1990; Cooke, 2006; Hwang et al., 2020; Szalma & Hancock, 2011). 그러나 임상 현장에서 주로 시행하는 순음 청력검사나 어음청각검사 결과를 통해 소음이 존재하는 일상생활 속 의사소통 능력을 평가 및 예측하는 데 한계가 있다(Kim & Lee, 2020; Taylor, 2003; Wilson, 2011). 따라서 청자의 소음 상황 속 어음인지능력을 측정하고 보청기나 인공와우 등의 청각보조기기 효과를 정량화하기 위해 다양한 소음 하 문장인지 검사도구의 개발 및 표준화 관련 연구가 진행되었다(Akeroyd

et al., 2015; Killion et al., 2004; Kollmeier et al., 2015; Nilsson et al., 1994). 청각보조기기의 착용효과를 정량화하기 위해서는 기본청력검사 시 사용하는 헤드폰 혹은 삼입이어폰 대신 라우드스피커를 사용하여 음장청각검사를 시행해야 한다(Ko & Lee, 2021). 음장청각검사를 시행할 때 음장검사의 환경이 결과에 유의한 영향을 미칠 수 있으므로 라우드스피커를 청자로부터 최소 1 m 이상의 거리에 두고, 자리에 앉은 청자의 머리 높이에 위치시키며, 기준점을 향하게 놓는 등 국제 표준에서 제시한 검사 환경에서 검사를 진행하는 것이 중요하다(British Society of Audiology, 2019; International Organization for Standardization, 2009). 적절한 음장검사 환경이 주어졌다면 검사의 목적에 따라 제시 레벨, 제시 방향, 제시 조건 등을 변경할 수 있다.

청각보조기기 착용 전후 소음 상황에서의 어음인지능력의 변화를 알아보기 위해 주로 사용하는 검사도구는 대부분 국외에서 개발된 검사로 Hearing in Noise Test (HINT; Nilsson

© 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.

et al., 1994), Quick Speech in Noise Test (Killion et al., 2004), Matrix Test (Kollmeier et al., 2015) 등이 있다. 위 검사 모두 소음 하 문장인지능력을 평가하는 변동형 검사도구로, HINT와 Matrix Test의 경우 약 65 dB SPL에서 소음을 제시한 채 청자가 제시한 문장을 50% 인지할 수 있는 신호대소음비(signal-to-noise ratio, SNR)를 결과로 산출한다(Lee & Yi, 2017). 위에 열거한 검사도구 중 HINT와 Matrix Test는 한국어 버전으로 개발되어 있어 임상 혹은 연구 목적으로 활용 가능하다. 한국어 HINT 검사(Moon et al., 2008)의 경우 목록당 20개의 일상생활 문장이 있으며, 총 12개 목록(총 240문장)을 포함한다. HINT 검사 시 정면(0°)과 좌우(±90°)에 위치한 음장 스피커를 활용하는 청력검사가 가능하나 고가의 장비를 사용해야 하는 문제가 있고, 현재 한국어 HINT 검사도구의 판매가 중지되어 임상적 활용면에서 제한적인 상황이다. Matrix 문장인지검사의 경우 전 세계에서 공통적으로 사용할 수 있도록 단어 및 문장 구성, 녹음 화자 선택 및 소음 제작, 문장 최적화, 검사 방법 및 타당도 검사 방법까지 최대한 모든 과정을 동일하게 하여(Akeroyd et al., 2015) 다국어 Matrix 문장 목록을 개발하였다. Matrix 문장은 10 × 5 행렬(Matrix)에 포함된 50개 단어 중 각 열에서 한 개의 단어씩(주어 1개, 형용사 1개, 명사 1개, 수사 1개, 동사 1개) 임의로 선택하여 문장이 구성되므로, 모든 문장이 동일한 문장 구조를 가지게 된다. 주어, 형용사, 명사, 수사, 동사 각각 10개의 단어 중 한 개의 단어씩 무작위 선택하므로 총 10<sup>5</sup>개의 서로 다른 문장이 사용 가능하며, 평가 시 대상자가 문장 내 다섯 개 단어를 추측하여 모두 맞추기가 어려워 반복 측정 시 혜택이 있다(Kollmeier et al., 2015).

한국어 Matrix 문장인지검사도구 역시 Akeroyd et al.(2015)이 제안한 권고사항대로 문장 목록을 구성하고 최적화(optimization)하였으며(Yi, 2016), 헤드폰과 Matrix 어음스펙트럼 소음을 통한 검사 결과의 신뢰도를 입증하였다(Kim et al., 2017; Kim & Lee, 2018; Kim & Lee, 2021). 그러나 청각보조기기 착용효과 측정을 위해 음장검사 환경에서 청각보조기기 착용 전후 반복적인 측정이 필요한데, 음장검사 환경에서 다양한 소음을 사용하여 측정 결과의 신뢰도를 확인하지 않았다. 모든 평가에서 측정 시 체계적 오차(systematic error)와 무작위 오차(random error)를 포함하는 측정 오차(measurement error)가 존재하는데 이러한 측정 오차는 반복 측정된 결과의 신뢰도에 영향을 줄 수 있다. 만약 청각보조기기의 착용효과 혹은 청능훈련 및 청각재활의 효과를 파악하기 위해 임상이 및 연구자들이 반복 측정을 시행하였다면 측정된 결과가 같지 않을 때 이것이 무작위 오차 때문에 발생한 것인지 일정하게 유지된 것인지를 신뢰도 연구 없이 판단하기가 어렵다. 검사 재검사 신뢰도를 확인하는 방법 중 하나로 측정치가 정량적일 때 급내상

관계수(intraclass correlation coefficient, ICC)를 구하여 상대적 신뢰도를 확인할 수 있다(Shrout & Fleiss, 1979). 보통 ICC가 클수록 신뢰도가 높다고 구분하는데, 보통 0.61~0.80 사이의 값을 보인다면 상당한 신뢰도를 보인다고 판단하며, 0.81~1.00 사이의 값을 보이면 거의 완벽한 신뢰도를 보인다고 한다(Landis & Koch, 1977). 그러나 ICC 값이 높을지라도 반복측정에 따른 학습효과로 인한 측정 오차가 여전히 발생할 가능성이 있으므로 절대적 신뢰도(측정 오차 신뢰도)인 표준측정 오차(standard error of measurement, SEM)와 최소감지 변화값(minimal detectable change, MDC)을 확인하는 경우도 있다(Lu et al., 2008).

본 연구에서는 청각보조기기 착용효과를 판단하기 위해 임상이 혹은 연구자들이 음장 소음 하 어음청각검사를 반복하여 시행해야 함을 고려하여, 건청인을 대상으로 다양한 소음을 사용하여 음장 Matrix 문장인지역치를 반복 측정하였다. 결과를 통하여 음장에서의 한국어 Matrix 검사도구의 상대적, 절대적 신뢰도를 확인하고, Bland & Altman plot을 이용하여 측정 결과 간 일치도를 알아보았다.

## MATERIALS AND METHODS

### 연구 대상

본 연구에 참여한 대상자는 정상 청력을 가진 성인 55명(남 13명, 여 42명)으로 평균 연령은 36.58 ± 9.27세(연령 범위: 24~55세)였다. 55명 중 35명(남 7명, 여 28명)의 대상자는 20~30대였고, 나머지 20명(남 7명, 여 13명)은 40~50대 대상자였다. 순음청력검사 결과 대상자의 양 귀 모두 250~8,000 Hz 이내 옥타브 단위 주파수에서 20 dB HL 미만의 순음청력역치를 보였으며, 65 dB SPL에서 어음을 제시한 결과 95% 이상의 단어인지도, 문장인지도를 보였다. 순음청력검사와 어음청각검사를 위해 청력검사기(AudioStar pro; Grason-Statler, Eden Prairie, MN, USA), 헤드폰(TDH-39P, Telephonics, Farmingdale, NY, USA)을 사용하였다. 대상자 모두 이경을 통한 육안 관찰 시 외이 및 고막에 이상이 없었으며, 임피던스 청력검사기(Resonance R25C; Resonance, Gazzaniga, Italy)를 이용하여 양 귀의 고막운동도검사를 실시한 결과 모두 A형 고막운동성을 보였다. 대상자 모두 이과 및 신경학적 병력이 없었으며, 연구에 참여하기 전 본 연구의 목적 및 절차에 대한 설명을 듣고 연구 참여에 동의하였다.

### 연구 절차

본 연구에서는 대상자의 정면에 위치한 라우드스피커를 이용한 음장검사 조건에서 한국어 Matrix 문장인지검사 결과의 검

사 재검사 신뢰도를 측정하고자 하였다. 음장 소음 하 어음청각 검사를 시행하기 위해 Oldenburg Measurement Applications (OMA) software (HörTech gGmbH, Oldenburg, Germany) 를 노트북(gram, LG electronics, Seoul, Korea)에 설치한 후, OMA software에서 제공되는 Matrix 문장과 소음을 Fireface UCX 디지털 아날로그 변환기(RME, Haimhausen, Germany)와 대상자로부터 1 m 거리, 귀 높이에 위치한 라우드스피커(SC-M53; DENON, Japan)를 통해 제시하였다. 모든 실험은 국제표준에서 권고하는 소음허용수준(International Organization for Standardization, 2012)을 만족하는 방음실에서 실시하였고, OMA 소프트웨어를 이용하여 음장검사 환경을 교정하였다. 검사에 사용한 Matrix 문장은 선행 연구에서 문장 내 단어의 레벨을 수정(level adjustment) 및 최적화하고(Yi, 2016) 목록 검증을 완료한 문장이었다(Kim & Lee, 2018). 다양한 소음 하 문장인지검사의 목적을 고려하여 본 연구에서는 Matrix 문장의 스펙트럼을 가진 어음스펙트럼과 강도 변동(amplitude modulation) 특성을 갖도록 제작된 International Collegium of Rehabilitative Audiology (ICRA) noise 5-250, International Speech Test Signal (ISTS) 소음을 사용하여 검사-재검사 신뢰도 평가를 진행하였다. ICRA 소음은 한명의 화자가 보통대화음레벨에서 녹음한 영어 어음을 필터링 한 후 표준 어음스펙트럼을 가지도록 제작한 소음이다. ICRA 소음은 어음의 장기평균어음스펙트럼과 어음의 envelope를 포함하지만 어음내 배음 구조(harmonic structure)를 가지고 있지 않은 광대역 소음이다(Dreschler et al., 2001). ISTS는 6개의 서로 다른 언어를 모국어로 사용하는 여성화자가 동일한 이야기("The north wind and the sun")를 읽도록 녹음한 후, 분할(segmentation)과 무작위 조합을 통해 개발한 표준테스트신호이다(Holube et al., 2010).

모든 소음 하 문장인지검사는 OMA 소프트웨어를 이용하여 진행하였고, SSN, ICRA, ISTS 소음의 레벨은 65 dB SPL에 고정하여 제시하였고 대상자의 반응에 따라 문장 레벨을 조절하는 변동형 검사로 진행하였다. 대상자가 문장 내 단어를 따라 말한 반응에 따라 최대우도측정(maximum likelihood estimator)에 근거하여(Brand & Kollmeier, 2002) 그 다음 자극음의 강도를 변동 조절하였다. 아래의 로지스틱 함수 Eq [1]을 통해 50% 소음 하 문장인지도에 해당하는 SNR을 구하여 speech reception threshold (SRT) 결과를 도출하였다.

$$p(L, SRT, s) = \frac{1}{1 + e^{4s(SRT-L)}} \quad [1]$$

S is the slope of the psychometric function, SRT is the

signal-to-noise ratio that yields 50% intelligibility, L is the signal to noise ratio.

검사-재검사 신뢰도 측정을 위해 0 dB SNR에서 소음하 문장인지검사를 시작하였고, Matrix Test의 훈련 효과를 최소화하기 위해 SSN 소음을 통해 친숙화를 위한 연습검사(practice)를 먼저 진행하였다. 연습검사가 종료된 후 SSN, ICRA, ISTS 소음을 통해 소음 하 문장인지도를 측정하였고, 순서에 따른 이월(carryover)효과를 고려하여 라틴스퀘어 실험 설계에 따라 검사를 진행하였다. 첫 번째 검사 진행 2주 후 SSN, ICRA, ISTS 소음을 사용하여 재검사를 진행하였다.

### 통계 분석

수집된 자료는 SPSS version 25.0 (IBM Corp., Armonk, NY, USA)을 이용하여 분석하였다. 검사-재검사 결과의 일치성(agreement)을 평가하기 위해 상대적 신뢰도지수인 ICC를 이용하였고, 이원배치변량(임의)효과모형(two-way random effects model)을 기준으로 ICC(2,1) 결과를 분석하였다. 개인 점수의 정확도를 결정하기 위해 절대적 신뢰도지수인 SEM과 MDC를 구하였다. SEM은 검사-재검사 결과의 표준편차(standard deviation, SD)와 ICC의 오차변량의 제곱근을 사용하여 구하였고(SEM = SD×√1-ICC), MDC는 95% 신뢰구간에서 MDC = 1.96×√2×SEM, MDC% = (MDC/mean × 100%)를 이용하여 산출하였다(Haley & Fragala-Pinkham, 2006). 검사-재검사의 두 검사 결과의 일치도를 시각적으로 확인하기 위해 Bland-Altman 도표(Bland & Altman, 1986; Giavarina, 2015)를 제시하였다. 검사-재검사 측정 결과 간 차이의 95%가 놓이는 범위에 대한 정보를 직관적으로 확인하기 위해 95% 일치한계수준(limits of agreement, LoA)을 확인하였다. 측정점수가 95% 신뢰구간(confidence interval, CI) 안에서 0점에 가까울수록 오차가 적고, 밖에 있으면 두 점수 차의 평균이 체계적 오차(systematic error)로 발생되고 있음을 의미한다. 추가로 피어슨 상관 분석으로 세 가지 소음을 통해 측정된 SRT 간 유의한 상관성이 있는지 확인하였다.

## RESULTS

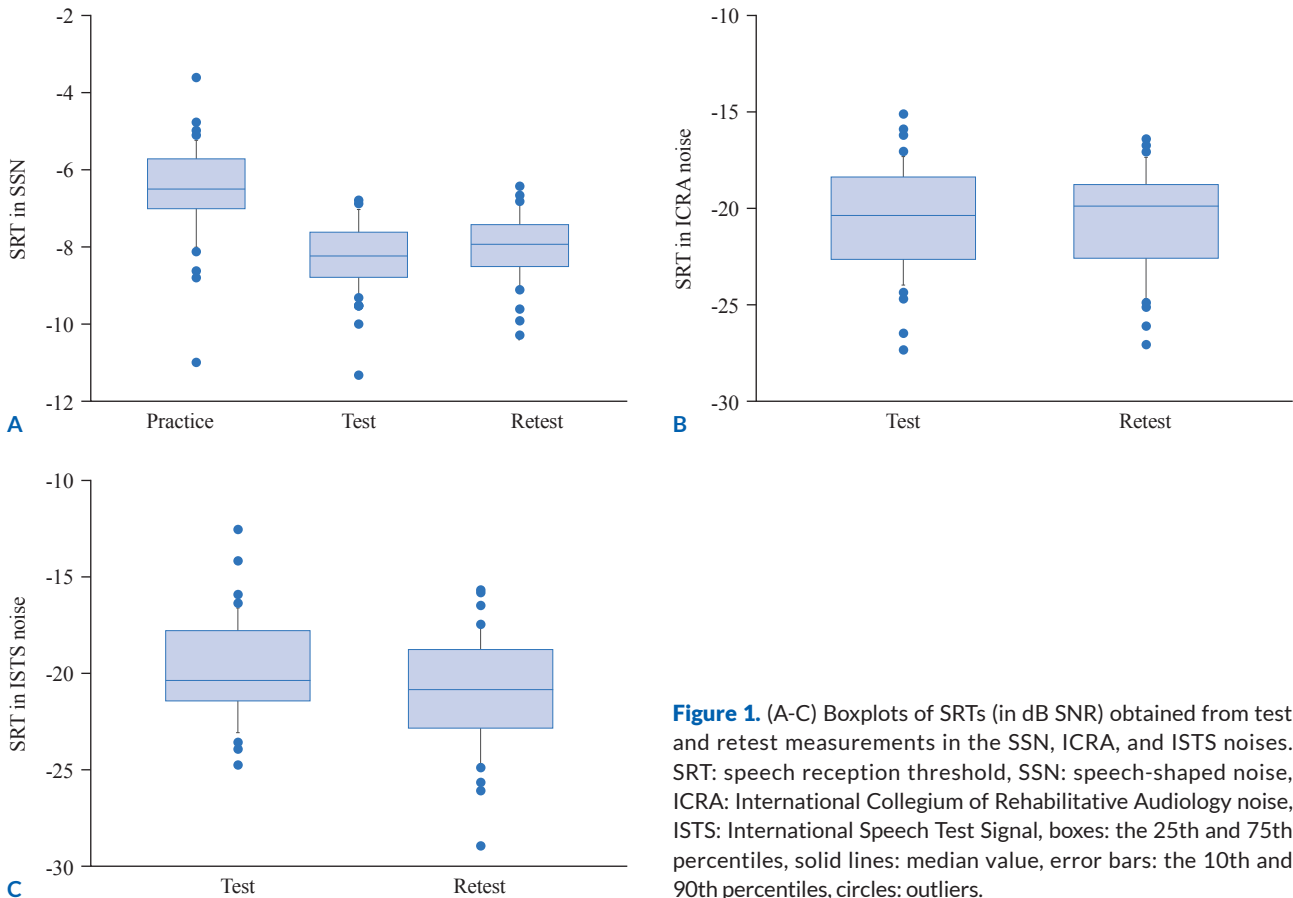
본 연구에서는 2주 간격을 두고 검사-재검사를 진행하였으며, 연습검사는 SSN을 통해서만 진행하였다. Figure 1A는 SSN을 통해 연습검사, 실제 검사-재검사에서 측정한 평균 SRT를 보여준다. Figure 1B는 ICRA 소음을 통해 구한 검사-재검사 평균 SRT를, Figure 1C는 ISTS 소음을 통해 구한 검사-재검사 평균 SRT를 보여준다. 먼저 비변동성 어음스펙트럼 소음인 SSN을 통해 연습 과정 시 측정한 평균 SRT는 -6.49 dB SNR (SD, 1.22;

range, -11.00 to -3.60 dB SNR)이었다. 연습 후 측정한 검사에서 측정한 평균 SRT는 -8.22 dB SNR (SD, 0.88; range, -11.30 to -6.80 dB SNR), 2주 후 측정한 재검사에서 측정한 평균 SRT는 -7.96 dB SNR (SD, 0.84; range, -10.30 to -6.40 dB SNR)이었다. 연습 과정 시 측정한 평균 SRT에 비해 검사, 재검사 평균 SRT가 1.5 dB 이상 더 낮아졌다(좋아졌다). 어음의 변동성 특징을 가진 ICRA 소음을 통해 측정한 검사-재검사 평균 SRT는 -20.69 dB SNR (SD, 2.75; range, -27.40 to -15.20 dB SNR), -20.67 dB SNR (SD, 2.71; range, -27.20 to -16.50 dB SNR)이었다. 마지막으로 ISTS 소음을 통해 측정한 검사-재검사 평균 SRT는 -19.88 dB SNR (SD, 2.56; range, -24.80 to -12.60 dB SNR), -20.92 dB SNR (SD, 2.81; range, -29.00 to -15.70 dB SNR)이었다.

Table 1은 세 가지 소음 각각에서 검사(test), 재검사(retest) 결과 간 일치도 분석 결과를 보여준다. SSN을 제시하였을 때 검사-재검사 신뢰도 ICC (95% CI)는 0.85 (0.75~0.92), ICRA 소음을 이용한 경우 ICC (95% CI)는 0.86 (0.75~0.93), ISTS 소음을 통해 측정한 ICC (95% CI)는 0.84 (0.71~0.91)로( $p < 0.001$ ) 유사한 안정성 신뢰도를 갖음을 확인하였다. 개인 점수의 정확

도를 결정하기 위해 절대적 신뢰도지수인 SEM과 MDC를 산출하였고, 이를 Table 1에 제시하였다. 여기서 MDC는 두 결과 간 점수의 변화가 일어났을 때 단순히 오류 때문에 생긴 변화가 아니라 대상자 개인에게서 생긴 변화라고 간주할 수 있는 최소한의 점수 변화량을 의미한다. SSN의 경우 SEM은 0.33, MDC는 0.92였고, ICRA 소음의 경우 SEM은 1.02, MDC는 2.83이었다. ISTS 소음의 경우 SEM은 1.07, MDC는 2.98로 SSN에 비해서 ICRA, ISTS 소음일수록 MDC 값이 비교적 더 큼을 확인하였다. SSN의 경우 MDC 값(MDC%)은 0.92 (11.38%), ICRA 소음의 경우 2.83 (13.68%), ISTS 소음의 경우 2.98 (14.54%)로 전반적으로 무작위측정오류(random measurement error)의 정도가 수용할만한 수준임을 확인하였다.

Figure 2는 SSN, ICRA, ISTS 소음을 제시하여 반복 측정한 음장 소음 하 문장인지역치 간 차이를 비교하기 위한 Bland & Altman plot이다. Figure 2를 통해 나타냈듯이, SSN의 경우 검사-재검사 일치한계(LoA)의 범위가 -1.46에서 +0.95, ICRA 소음의 경우 LoA 범위가 -3.7에서 +3.70, ISTS 소음의 경우 LoA 범위가 -3.01에서 +4.70이었다. 검사-재검사에서 추정된 두 결과 값이 실제 본 연구에 참여한 대상자 중 1~2명을 제외하고는



**Figure 1.** (A-C) Boxplots of SRTs (in dB SNR) obtained from test and retest measurements in the SSN, ICRA, and ISTS noises. SRT: speech reception threshold, SSN: speech-shaped noise, ICRA: International Collegium of Rehabilitative Audiology noise, ISTS: International Speech Test Signal, boxes: the 25th and 75th percentiles, solid lines: median value, error bars: the 10th and 90th percentiles, circles: outliers.



측정 가능한 결과 범위 내에 대부분 분포하고 있으므로 신뢰할 수 있는 수준임을 확인하였다.

마지막으로, 피어슨 상관 분석으로 SSN, ICRA, ISTS 소음을 통해 측정한 SRT 간 유의한 상관성이 있는지 확인하였다. 분석 결과, SSN을 통해 구한 SRT가 높을수록 ICRA 소음을 제시하고 측정한 SRT ( $r = 0.70, p < 0.001$ ) 혹은 ISTS 소음을 제시하고 측정한 SRT ( $r = 0.62, p < 0.001$ ) 결과가 유의하게 높았다. ICRA 소음 하에서 구한 SRT가 높을수록 ISTS 소음 하에서 구한 SRT ( $r = 0.73, p < 0.001$ ) 결과가 유의하게 높았다.

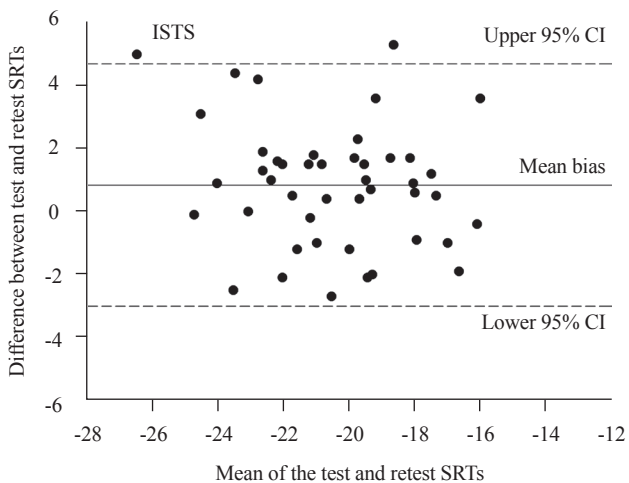
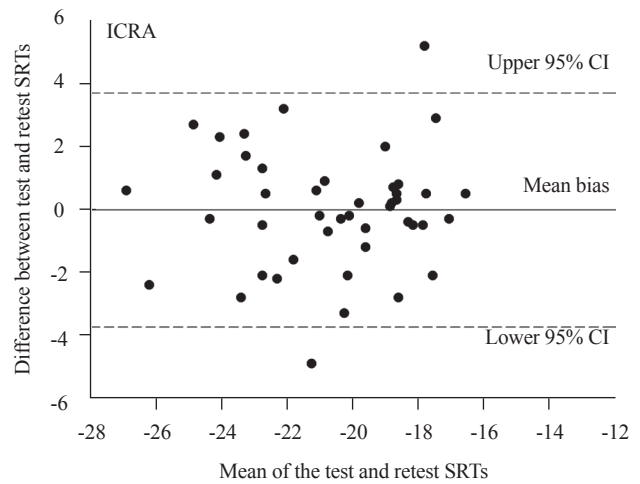
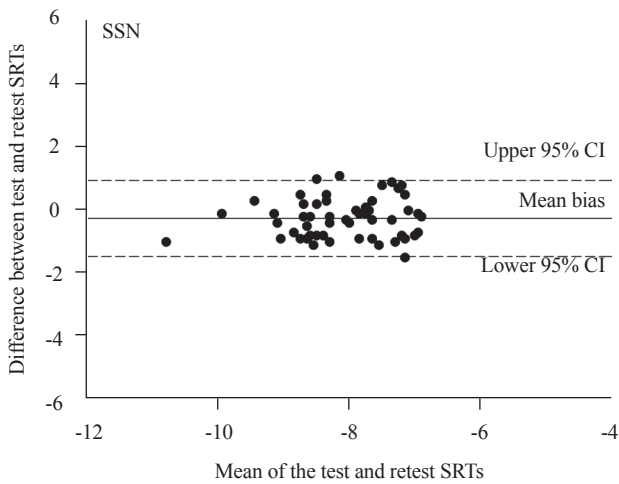
## DISCUSSIONS

난청인의 소음 하 의사소통능력을 신뢰도 있게 정량화하기 위해서 소음 하 어음인지검사도구의 신뢰도, 타당도 등을 검증하는 것이 중요하다. 소음 상황에서 난청인의 청각보조기기 착용효과를 평가하기 위해서는 헤드폰이 아닌 음장검사 조건에서 검사를 시행하여야 하므로 이러한 검사 조건에서 결과의 신뢰도를 확인하는 것이 필요하다. 국내에서 사용 가능한 변동형 소음 하 어음인지검사도구의 신뢰도 검증이 부족한 점을 고려하여 본 연구

**Table 1.** Test-retest reliability of the Korean Matrix sentence-in-noise test in the sound-field testing condition

Noise type	Test (dB SNR)	Retest (dB SNR)	ICC <sub>2,1</sub> (95% CI)	SEM	MDC (MDC%)
SSN	-8.22 ± 0.88	-7.96 ± 0.84	0.85 (0.75~0.92)	0.33	0.92 (11.38%)
ICRA	-20.69 ± 2.75	-20.67 ± 2.71	0.86 (0.75~0.93)	1.02	2.83 (13.68%)
ISTS	-19.88 ± 2.56	-20.92 ± 2.81	0.84 (0.71~0.91)	1.07	2.98 (14.54%)

Values are presented as mean ± standard deviation or number unless otherwise indicated. ICC<sub>2,1</sub>: intraclass correlation coefficients, model (2, 1), CI: confidence interval, SEM: standard error of measurement, MDC: minimal detectable change, SSN: speech-shaped noise, ICRA: International Collegium of Rehabilitative Audiology noise, ISTS: International Speech Test Signal



**Figure 2.** Bland-Altman plots showing the difference between the test and retest measurements against the mean of the two measurements. SRTs: speech reception thresholds, CI: confidence interval, SSN: speech-shaped noise, ICRA: International Collegium of Rehabilitative Audiology noise, ISTS: International Speech Test Signal, dotted lines: the upper and lower limits of agreement for 95% CI, solid line: the mean of the difference (bias).

에서는 음장 Matrix 문장인지검사를 반복 시행하여 검사-재검사 일치율 및 신뢰도를 확인하고자 하였다.

본 연구에서는 정면에 위치한 라우드스피커를 통해 SSN, ICRA, ISTS 소음을 제시하고 변동형 Matrix 소음 하 문장인지 검사 결과의 신뢰도를 분석하였다. 먼저 SSN 소음을 통해 검사를 시행한 결과 검사-재검사 건청 성인의 SRT를 평균하면 -8.09 dB SNR이었다. 선행 연구에서 헤드폰과 SSN 소음을 통해 동일 레벨에서 검사를 시행하였을 때(Kim & Lee, 2021) 건청 성인의 평균 SRT가 -8.13 dB SNR이었으므로 본 연구와 선행 연구 검사 결과가 매우 유사함을 확인하였다. ICRA, ISTS 소음 모두 표준 장기평균어음스펙트럼 특징과 어음 envelope을 통한 변동 특징은 가지고 있지만 의미를 가지고 있지 않은 비의미 소음(non-meaningful noise)에 속한다. ICRA 혹은 ISTS 소음을 통해 검사를 시행하였을 때 검사-재검사 건청 성인의 평균 SRT가 약 -20 dB SNR이었다. 이는 65 dB SPL에서 제시된 소음 속에서 45 dB SPL에서 제시된 문장을 50% 가량 인지할 수 있었다는 것을 의미한다. SSN을 통해 측정한 검사 결과에 비해 ICRA 혹은 ISTS 소음에서 결과 측정 시 평균 SRT 값이 약 -20 dB SNR로 12 dB 만큼 SRT가 낮아졌고 표준편차 값은 커짐을 알 수 있다. 이는 다수의 선행 연구(Cooke, 2006; Fogerty et al., 2018; Miller et al., 2018)에서 보고한 바와 같이 건청 성인은 변동 소음의 강도가 일시적으로 낮은 진폭을 나타내는 순간 신호대소음비가 상승하여 목표 어음을 좀 더 많이 인지할 수 있었던 것이라고 볼 수 있다. 이러한 현상을 glimpsing, dip listening 혹은 fluctuating masker benefit이라고 일컫기도 한다(Cooke, 2006; Jensen & Bernstein, 2019). SSN 소음에 비해 ICRA 혹은 ISTS 소음을 사용하였을 때 표준편차가 컸던 이유는 본 연구 결과로 측정한 값이 50% 가량 인지할 수 있는 SNR이었으므로 소음에 비해 약 20 dB 가량 더 낮은 강도의 문장을 알아듣는 듣기 조건에서 소음의 무작위적 변동 특성에 따라 개인 간 차이가 비교적 크게 나타났을 것으로 생각된다.

본 연구에서는 검사-재검사 결과의 일치성(agreement)을 평가하기 위해 상대적 신뢰도지수인 ICC를 확인하였다. ICC가 클수록 높은 신뢰도를 의미하며 1은 완벽한 일치도를 0은 일치도가 전혀 없는 것으로 판단한다. Landis and Koch(1977)의 분류에 의하면 0.81~1 이내의 ICC 계수는 신뢰도가 매우 우수함을 의미한다. 본 연구에서는 소음의 종류에 상관없이 검사-재검사 일치도 분석 결과 ICC가 0.84~0.86으로 우수한 신뢰도를 보였다. 이러한 우수한 신뢰도를 보인 이유는 Matrix 검사도구 개발 절차 시 문장을 포함하는 단어 하나씩 최적화를 거쳐 비교적 우수한 목록 간 음향적 동질성을 가지기 때문이고, 소프트웨어를 이용한 변동형 검사 절차 및 레벨 교정으로 검사자 및 검사기에 관한 측정 오차를 비교적 줄일 수 있었을 것으로 생각한다.

평가도구의 신뢰도 측정이 중요한 이유는 임상에서 대상자를 반복 평가하였을 때 평가 결과가 일정하게 유지되는지, 측정된 결과의 변화가 측정 오차 때문에 발생한 것은 아닌지 확인이 필요하기 때문이다. 예를 들면 난청인을 대상으로 청능훈련 전후 소음 하 어음인지 평가 결과 훈련 전 SRT가 0 dB SNR에서 훈련 후 SRT가 -3 dB SNR로 낮아졌다면 이것이 측정 오차에 의한 것인지 실제 훈련에 의한 것인지 확인하지 않고는 훈련의 효과라고 말하기 어렵다. ICC 결과는 평가를 반복 시행하였을 때 평가 결과 간 일관성을 확인하는 목적으로 상대적 신뢰도를 확인하는데 유용하므로, 절대적 신뢰도지수인 SEM, MDC 값을 구하여 반복된 평가 시 우연히 발생 가능한 측정 오차를 확인하는 것이 중요하다. 본 연구에서는 세 가지 소음 종류, SSN, ICRA, ISTS를 사용하여 검사-재검사 측정을 시행한 후 상대적, 절대적 신뢰도지수를 확인하였다. 세 가지 소음 중 SSN을 제시하였을 때 SEM은 0.33, MDC는 0.92였으나, 변동형 특징을 가지는 ICRA 혹은 ISTS 소음을 제시하였을 때 SEM은 약 1.05, MDC는 약 2.9였다. SEM은 보통 총 평균값의 10% 미만, MDC는 측정값 중 최고 점수의 20% 미만인 경우 측정 오차가 작아 수용할만하다고 간주하므로(Lu et al., 2008) 본 연구에서 사용한 SSN, ICRA, ISTS 소음을 이용하여 음장 소음 하 어음청각검사를 이용할 경우 측정 오차가 수용할 만한 수준으로 상대적, 절대적 신뢰도가 적절하다고 볼 수 있다.

본 연구의 제한점으로는 검사-재검사 기간을 2주로 하였다는 점이다. 추후 장기간 동안 여러 번에 걸쳐 반복 측정하여 검사-재검사 신뢰도를 확인하는 것이 필요하겠다. 본 연구에서는 난청인 검증 전, 건청인을 대상으로 신뢰도 검증을 진행하는 것이 목적이었다. 후속 연구에서 난청인을 대상으로 음장검사 결과의 신뢰도 및 타당도를 검증하는 것이 필요하겠다. 마지막으로, 추후 연구에서는 다양한 음장검사 환경 속에서 검사의 타당도를 확인하는 것이 필요하겠다.

본 연구에서 건청인을 대상으로 음장 Matrix 문장인지검사를 반복 시행한 결과 검사-재검사 일치율 및 신뢰도가 적절함을 확인하였다. 세 가지 소음 중 SSN의 경우 다른 소음 종류에 비해 무작위 오차의 정도가 가장 작았지만, 목적에 따라 ICRA, ISTS 소음도 수용할 만한 소음 종류임을 확인하였다. 신뢰도, 타당도가 입증된 변동형 소음 하 음장검사도구가 부족한 현 시점에서 본 연구의 결과를 소음 하 평가 및 재활 등 다양한 임상 및 연구 목적으로 유용하게 사용할 수 있을 것이다.

**중심 단어:** 검사 재검사 신뢰도, 한국어 Matrix 문장검사, 음장 소음 하 문장인지검사도구의 검증.

### Ethical Statement

This study was approved by the Institutional Review Board of Hallym University of Graduate Studies (#IRB: HUGSAUD256198).

### 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).

### ORCID iD

Yukyeong Jung <https://orcid.org/0000-0002-9774-9899>  
 Jae Ho Han <https://orcid.org/0000-0001-8364-0780>  
 Seokyu Choi <https://orcid.org/0000-0002-2718-6832>  
 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.
- Bland, J. M. & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 1(8476)*, 307-310.
- 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.
- British Society of Audiology. (2019, April). *Practice Guidance: The Acoustics of Sound Field Audiometry in Clinical Audiological Applications*. British Society of Audiology. Retrieved from <https://www.thebsa.org.uk/resources/guidelines-acoustics-sound-field-audiometry-clinical-audiological-applications/>.
- Bronkhorst, A. W. & Plomp, R. (1990). A clinical test for the assessment of binaural speech perception in noise. *Audiology, 29(5)*, 275-285.
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America, 119(3)*, 1562-1573.
- 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.
- Fogerty, D., Carter, B. L., & Healy, E. W. (2018). Glimpsing speech in temporally and spectro-temporally modulated noise. *The Journal of the Acoustical Society of America, 143(5)*, 3047-3057.
- 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.
- 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.
- Hwang, J. S., Jung, Y., & Lee, J. H. (2020). The effect of background noise on immediate free recall of words in younger and older listeners with hearing loss. *Journal of the American Academy of Audiology, 31(10)*, 701-707.
- International Organization for Standardization. (2009). *Acoustics: Audiometric Test Methods. Part 2: Sound Field Audiometry with Puretone and Narrow-Band Test Signals (ISO 8253-2:2009)*. Geneva: International Organization for Standardization.
- 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.
- 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. 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.
- Kim, K. H. & Lee, J. H. (2021). Influence of measurement procedure on the Korean Matrix sentence-in-noise intelligibility for normal-hearing listeners. *Korean Journal of Otorhinolaryngology-Head and Neck Surgery, 64(1)*, 7-12.
- 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, N. K., Yi, D., Bahng, J., & Lee, J. H. (2017). Study on the reliability and the validity for the sentence-in-noise scores at the fixed signal-to-noise ratios. *Audiology and Speech Research, 13(1)*, 62-69.
- Ko, Y. & Lee, J. H. (2021). Clinical importance of sound-field speech recognition test for measurement of hearing aid benefit. *Korean Journal of Otorhinolaryngology-Head and Neck Surgery, 64(5)*, 310-320.
- Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, 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.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33(1)*, 159-174.
- Lee, J. H. & Yi, D. W. (2017). A comparison of adaptive sentence-in-noise tests. *Audiology and Speech Research, 13(1)*, 9-18.
- Lu, W. S., Wang, C. H., Lin, J. H., Sheu, C. F., & Hsieh, C. L. (2008). The minimal detectable change of the simplified stroke rehabilitation

- assessment of movement measure. *Journal of Rehabilitation Medicine*, 40(8), 615-619.
- Miller, R. E., Gibbs, B. E., 2nd, & Fogerty, D. (2018). Glimpsing speech interrupted by speech-modulated noise. *The Journal of the Acoustical Society of America*, 143(5), 3058-3067.
- 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.
- Taylor, B. (2003). Speech-in-noise tests: How and why to include them in your basic test battery. *The Hearing Journal*, 56(1), 40, 42-46.
- Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428.
- Szalma, J. L. & Hancock, P. A. (2011). Noise effects on human performance: A meta-analytic synthesis. *Psychological Bulletin*, 137(4), 682-707.
- 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.