A Systematic Review and Meta-Analysis for Better Measurement of Listening Effort in Adults with Hearing Loss

Chanbeom Kwak, Woojae Han

1Laboratory of Hearing and Technology, 2Division of Speech Pathology and Audiology, 3Research Institute of Audiology and Speech Pathology, College of Natural Sciences, Hallym University, Chuncheon, Korea

The present study aimed to identify better tools for measuring listening effort in the hearing-impaired and/or hearing aid users and to suggest its clinical implication by using systematic review and meta-analysis. To search articles from six electronic databases, ‘listening effort’, ‘hearing loss/hearing impaired’, and ‘hearing aid(s)’ were used as their key terms. Although 8,761 articles were found initially, only 19 articles which met the inclusion criteria were applied to the further review and analysis. In a checklist of the study quality, there was no significant difference between the articles. Based on meta-analysis of three subgroups (i.e., objective, subjective, and self-report rating), the objective measurements showed statistically significant and the highest effect size (1.031, 95% confidence interval: -0.106–2.169), whereas the self-report rating had the lowest effect size (-1.280, 95% confidence interval: -4.180–1.620) with no significance. Although the funnel plot was asymmetrical, the Egger’s regression asymmetry test revealed no publication bias. In sum, the objective measurement showed the most effective way to evaluate the listening effort for the hearing-impaired and/or hearing aid users. However, there was a non-negligible variance between the types of measurements. In the following study, we suggest to investigate certain relationship between listening effort and speech perception performance in the hearing-impaired and/or hearing aid users and to establish standardized criteria for clinical purposes of the listening effort.

Key Words: Hearing aid, Hearing loss, Listening effort, Objective measure, Self-report rating, Subjective measure.

INTRODUCTION

Listening effort is a construct that refers to the cognitive resources required for understanding speech in difficult listening conditions. It is a psychological construct that refers to the cognitive resources required for understanding speech in difficult listening conditions. The present study aimed to identify better tools for measuring listening effort in the hearing-impaired and/or hearing aid users and to suggest its clinical implication by using systematic review and meta-analysis. To search articles from six electronic databases, ‘listening effort’, ‘hearing loss/hearing impaired’, and ‘hearing aid(s)’ were used as their key terms. Although 8,761 articles were found initially, only 19 articles which met the inclusion criteria were applied to the further review and analysis. In a checklist of the study quality, there was no significant difference between the articles. Based on meta-analysis of three subgroups (i.e., objective, subjective, and self-report rating), the objective measurements showed statistically significant and the highest effect size (1.031, 95% confidence interval: -0.106–2.169), whereas the self-report rating had the lowest effect size (-1.280, 95% confidence interval: -4.180–1.620) with no significance. Although the funnel plot was asymmetrical, the Egger’s regression asymmetry test revealed no publication bias. In sum, the objective measurement showed the most effective way to evaluate the listening effort for the hearing-impaired and/or hearing aid users. However, there was a non-negligible variance between the types of measurements. In the following study, we suggest to investigate certain relationship between listening effort and speech perception performance in the hearing-impaired and/or hearing aid users and to establish standardized criteria for clinical purposes of the listening effort.

Key Words: Hearing aid, Hearing loss, Listening effort, Objective measure, Self-report rating, Subjective measure.
2016: Wingfield et al., 2006) 균형적으로 청사는 그러한 청취 상황에서 쉽게 피로하거나 집중을 포기하게 된다.

여러 연구들에서 이미 밝혀졌지만, 난청인들은 배경 소음이나 반향음 등 청취에 어려운 환경에서 저조한 어음인지를 보이며 (Ayasse et al., 2017; Ayasse & Wingfield, 2018; Desjardins & Doherty, 2013: Huber et al., 2018; Krueger et al., 2017; Picou & Ricketts, 2018; Ricketts et al., 2019; Ward et al., 2017; Ze- kveld et al., 2009). 보청기나 인공외래 등과 같은 청각보조기 기능을 활용하여 청력 저하의 보완은 물론 저조한 어음인지도의 향상을 기대한다(Alhanbali et al., 2018; Alhanbali et al., 2019; Desjardins & Doherty, 2013; Giroud et al., 2017; Krueger et al., 2017; Mackersie et al., 2015; Picou & Ricketts, 2018; Ricketts et al., 2019; Zekveld et al., 2009). 최근 학계에서는 이러한 난청인들의 어음인지력 향상을 위해 요구하는 인지적 자원인 청취 능력을 제시하고 있다. Fichora-Fuller et al.(2016)은 Framework for Understanding Effortful Listening (FUEL)을 설명하면서 기존 연구들은 보청기 작용을 통한 난청인들의 가정적 보완 및 더 나은 어음인지력의 향상이 주된 목적이었으나, 이를 향상시키기 위해 오히려 난청인들에게 요구되는 과도한 청취 능력을 바람직하지 않다고 주장하였다. 난청 노인들은 대상으로 객관적 측정 방법론 동공 확장하기 사용하여 언어적 복잡성에 따른 문장 인지와 청취 능력에 대해 보고한 Ayasse & Wingfield(2018)의 연구에서도 난청 노인은 청각 능력에 비해 감각의 난이도가 어려워짐에 따라서 동공의 확장을 유의미하게 발전하였다. 다시 말해, 동일한 언어 능력을 보이며라도 난청으로 인한 문장 인지의 저조한 수행력과 더불어 과제의 복잡성으로 유발되는 과도한 인지적 자원은 난청인에게 더 높은 청취 능력을 요구하게 됨을 알 수 있다. 그러나 난청 어음인지력은 수행하기 위해 수반되는 과도한 인지적 노력을 인해 난청인들은 쉽게 피로감을 느끼게 되고 어려운 청취 환경을 겪는 등 장기적 관점에서는 난청인의 수행력에 부정적 인 영향을 미치게 된다. 어음인지능과 청각 능력은 보청기로 측정한 청취 능력 간 연관성에 대해 연구한 Krueger et al.(2017)의 결과 역시, 배경 소음의 정도가 증가함에 따라 장애물에 난청의 어음인지능에 대한 수용성이 그룹에 비해 급격하게 감소할 뿐만 아니라, 이에 요구되는 청취 능력 역시 난청 그룹은 유의미하게 더 높았다. 또한, 연구에 참여한 어음인지능의 난청인들에게 보청기 작용을 통해 일차적으로 어음인지를 향상시켰다. 보청기의 청취 능력은 인지 기능의 복합적인 과정으로 인해 보청기 작용이 향상시켰고, 보청기 작용을 통해 일차적으로 어음인지를 향상시켰다. 보청기의 청취 능력은 인지 기능의 복합적인 과정으로 인해 보청기 작용이 향상시켰다. 보청기의 청취 능력은 인지 기능의 복합적인 과정으로 인해 보청기 작용이 향상시켰다. 보청기의 청취 능력은 인지 기능의 복합적인 과정으로 인해 보청기 작용이 향상시켰다. 보청기의 청취 능력은 인지 기능의 복합적인 과정으로 인해 보청기 작용이 향상시켰다.
**Table 1.** Inclusion criteria based on the Participants, Intervention, Control, Outcomes, and Study designs (PICOS) strategy

| PICOS       | Contents                                                                 |
|-------------|---------------------------------------------------------------------------|
| Participants| Adults (18+ years) with hearing loss and/or hearing aid                    |
| Intervention| Objective, subjective, and self-report rating measurements                 |
| Control     | Adults (18+ years) with normal hearing or repeated measures               |
| Outcomes    | Listening effort expressed by objective (i.e., pupil dilation), subjective, (i.e., accuracy), and rating scale (i.e., Likert scale) |
| Study designs| Randomized controlled trials, non-randomized controlled trials, cohort studies, between-group comparison, and repeated measures with additional purposes |

**Figure 1.** Flowchart of the study search and selection process based on PICOS criteria. PICOS: Participants, Intervention, Control, Outcomes, Study design, ERP: event-related potential, EEG: electroencephalography, CTMT: Comprehensive Trail-Making Test.
ASR Methods for Listening Effort Measurement

Table 2. Scientific study validity criteria based on CAMARADES checklists (PROSPERO, 2014)

| Article                        | Randomization | Controls | Sample size calculation | Publication after peer review | Outcome measure | Statement of potential conflict of interest | Study quality score |
|-------------------------------|---------------|----------|-------------------------|-------------------------------|-----------------|------------------------------------------|---------------------|
| Ayasse & Wingfield, 2018      | 0             | 1        | 0                       | 1                             | 1               | 1                                        | 4                   |
| Desjardins & Doherty, 2013    | 0             | 1        | 1                       | 1                             | 1               | 1                                        | 5                   |
| Ward et al., 2017             | 0             | 1        | 0                       | 1                             | 1               | 1                                        | 4                   |
| Tun et al., 2009              | 0             | 1        | 0                       | 1                             | 1               | 0                                        | 3                   |
| Ayasse et al., 2017           | 0             | 1        | 1                       | 1                             | 1               | 1                                        | 4                   |
| Rosemann & Thiel, 2020        | 0             | 1        | 0                       | 1                             | 1               | 1                                        | 4                   |
| Bertoli & Bodmer, 2014        | 1             | 1        | 0                       | 1                             | 1               | 0                                        | 4                   |
| Giroud et al., 2017           | 1             | 1        | 1                       | 1                             | 1               | 0                                        | 4                   |
| Bernarding et al., 2013       | 1             | 1        | 0                       | 1                             | 1               | 0                                        | 4                   |
| Zekveld et al., 2009          | 1             | 1        | 1                       | 1                             | 1               | 0                                        | 5                   |
| Rosemann & Thiel, 2019        | 0             | 1        | 0                       | 1                             | 1               | 0                                        | 4                   |
| Mackersie et al., 2015        | 1             | 1        | 1                       | 1                             | 1               | 1                                        | 5                   |
| Alhanbali et al., 2019        | 1             | 1        | 0                       | 1                             | 1               | 1                                        | 5                   |
| Krueger et al., 2017          | 1             | 1        | 0                       | 1                             | 1               | 0                                        | 4                   |
| Holube et al., 2016           | 1             | 1        | 0                       | 1                             | 1               | 1                                        | 5                   |
| Picou & Ricketts, 2018        | 1             | 0        | 0                       | 1                             | 1               | 1                                        | 4                   |
| Huber et al., 2018            | 1             | 1        | 0                       | 1                             | 1               | 1                                        | 5                   |
| Ricketts et al., 2019         | 1             | 0        | 0                       | 1                             | 1               | 1                                        | 4                   |
| Alhanbali et al., 2018        | 0             | 0        | 1                       | 1                             | 1               | 1                                        | 4                   |

1 and 0 refer to “Yes” and “No”, respectively
Meta-analysis

본 연구에 포함된 총 35편의 문헌에서 추출 및 합성된 데이터가 메타 분석에 적합 여부를 확인 후, Comprehensive Meta-Analysis (Ver. 3, Biostat Inc., Englewood, NJ, USA)를 사용하여 분석을 실시하였습니다. 메타 분석은 주술 및 합성된 데이터의 특성에 따라 효과 크기(effect size)를 계산하는데 사용되는 측도가 결정된다(Kang, 2015). 본 연구에 포함된 문헌 35건의 데이터 중 20건은 실험군과 대조군의 평균과 표준편차를, 나머지 15건은 상관관계 결과를 보고하였기에 각 문헌의 평균, 표준편차, 상관관계 결과를 바탕으로 청취 노력에 대한 효과 크기를 추정하였습니다.

메타 분석의 특성 상 대다수의 문헌들이 청취 노력을 측정하기 위해 사용한 측정 방법이 상이하기 때문에, 표준화된 평균 차이(standardized mean difference, SMD)를 적용하여 효과 크기를 추정하고 변량 효과 모형의 95% 신뢰구간(confidence interval, CI)에서 검증하였습니다. 관측자 그림(funnel plot)의 산점도와 Egger's regression asymmetry test의 유의 확률 (p < 0.05)을 통해 출판 편향을 확인하였습니다. 또한 Cochrane's Q test와 Higgins I²-statistics를 근거로 각 문헌 간 분산의 이질 성(heterogeneity)를 확인하였습니다. Cochrane's Q test는 95%의 유의수준 (p < 0.05)을 기준으로 이질성의 여부를 판단하며, Higgins I² statistics는 0%에서 100%의 값으로 이질성에 대한 양적 기준을 제시한다. I²가 0%에서 25%는 낮은 수준의 이질성, 25%에서 75%는 중간 수준의 이질성, 75%에서 100%는 높은 수준의 이질성을 의미한다(Higgins et al., 2003).

RESULTS

Scores of study quality

CAMARADES checklist (PROSPERO, 2014)를 기준으로 평가한 문헌의 질적 평가와 연구의 임재적 편향은 통계/데이터마 이닝 및 그래프를 위한 언어인 R 프로그램(R Core Team, 2018)을 사용하여 카이 제곱 검정(chi-square test)을 수행하였습니다.

질적 평가에 대한 전체 문헌들의 평균은 4.20 (standard deviation: 0.62, range: 3~5)이었고, 문헌의 질적 평가에 대한 적합도 확인을 위한 카이 제곱 검정 결과는 문헌의 질적 평가 각 유의미한 차이를 보이지 않았다 ($\chi^2 = 3.7536, df = 19, p > 0.05$).

Characteristic analysis of studies in meta-analysis

Table 3은 PICOS 기준에 따라 각 문헌의 결과를 요약하여 정리하였습니다. 대상자는 난청 혹은 보청기를 착용한 18세 이상의 성인이었으며, 19편의 문헌에는 난청을 가진 노인(Ayassee et al., 2017; Ayasse & Wingfield 2018; Huber et al., 2018; Tun et al., 2019), 보청기를 착용한 노인(Giroud et al., 2017), 난청 노인과 보청기 착용 노인(Alhanbali et al., 2018), 난청을 가진 장년(Bernardining et al., 2013), 난청 장년과 보청기 착용 장년(Zekveld et al., 2009), 난청을 가진 장년과 노인(Bertoli & Bodmer, 2014; Holube et al., 2016; Rosemann & Thiel, 2019; Rosemann & Thiel, 2020; Ward et al., 2017), 보청기를 착용한 장년과 노인(Desjardins & Doherty, 2013; Krueger et al., 2017; Ricketts et al., 2019), 난청 장년 및 노인과 보청기 착용 장년 및 노인(Alhanbali et al., 2019; Picou & Ricketts, 2018), 난청 장년 및 노인과 보청기 착용 장년 및 노인(Mackersie et al., 2015)이 포함되었다.

측정 방법에 따른 분류에 있어서 객관적 측정을 수행한 14건의 문헌 중에는 동공 확장(Alhanbali et al., 2019; Ayasse et al., 2017; Ayasse & Wingfield, 2018) 및 연구 응직임의 정확도(Tun et al., 2009), EEG (Alhanbali et al., 2019) 및 사건 유발 전위(event-related potential, ERP)에 포함되는 P50-N1-P2 complex (Giroud et al., 2017), N1반응(Bernardining et al., 2013), P3반응(Bertoli & Bodmer, 2014), 피부 전도성(Alhanbali et al., 2019; Mackersie et al., 2015), 호흡수 및 심박수(Mackersie et al., 2015)가 포함되었다. 한편, 주관적 측정을 수행한 11건의 문헌 중에는 정확도(Desjardins & Doherty, 2013; Ward et al., 2017, Picou & Ricketts, 2018; Ricketts et al., 2019), 반응 시간(Alhanbali et al., 2019; Ricketts et al., 2019; Ward et al., 2017), 응답 시간(Picou & Ricketts, 2018) 및 작업 기억의 능력을 평가하는 회상 정확도(Tun et al., 2009), 종합 기호 점수(Rosemann & Thiel, 2020) 단어 회상 정확도(Bertoli & Bodmer, 2014)가 포함되었다. 마지막으로 자녀 보고 측정을 수행한 10건의 문헌들은 모두 척도의 기반인 설문 문항을 사용하였으며, 10점 척도(Alhanbali et al., 2019; Rosemann & Thiel, 2019; Rosemann & Thiel, 2020; Zekveld et al., 2009)가 가장 많이 적용되었다. 이어서 7점 척도(Holube et al., 2016; Huber et al., 2018; Krueger at al., 2017, 11점 척도(Ricketts et al., 2019), 20점 척도(Alhanbali et al., 2019), 100점 척도(Desjardins & Doherty, 2013)의 순으로 다양한 척도가 사용되었다.

Overall effectiveness of measurements for listening effort

변량 효과 모형을 사용한 전체 문헌의 효과 크기 결과는 Figure 2A에 제시하였습니다. 전체 문헌의 표준화된 평균 차이는 1.031(95% CI: -0.106~2.169)을 보였다. 청취 노력의 측정 방법에 따른 효과를 확인하기 위해 세 가지 측정 방법에 따른 하위군 분석을 수행하였으며, 객관적 측정에 따른 청취 노력의 결과는 1.686의 효과 크기와 통계적 유의미성(0.95 CI: 1.233~
### Table 3. Summary of the extracted and synthesized data from 19 included articles

| Study                                                                 | Participants                                      | Test materials                                                                 | Study design                                      | Outcome measures                        | Main findings                                                                 |
|----------------------------------------------------------------------|---------------------------------------------------|--------------------------------------------------------------------------------|---------------------------------------------------|-----------------------------------------|--------------------------------------------------------------------------------|
| Ayasse & Wingfield, 2018                                             | 14 older adults with a mild-to-moderate hearing loss | IEEE sentences corpus with 25 dB SL                                               | Between group comparison, repeated measures       | Peak pupil dilation                     | The measurement of pupil dilation as an index of processing effort showed effort to increase with task difficulty until a difficulty tipping point was reached. |
| Desjardins & Doherty, 2013                                           | 16 old adults with hearing aid (aged 59–76 years, mean: 66.86, SD: 6.7) | R-SPIN test for both high- and low-context context presented by female talker    | Between group comparison, repeated measures       | Dual-task paradigm (DPRT)               | Both ONH and OHI had significantly (p < 0.001) greater differences in performance scores between the high- and low-context conditions compared with the YNH. However, there was no significant (p > 0.05) difference in scores between the ONH and OHI. This suggests that the older participants benefited more from the contextual cues in the sentences than the younger participants. |
| Ward et al., 2017                                                    | 21 older adults who showed hearing thresholds of ≤ 25 dB at 0.25 to 2 kHz and 45 ≤ dB at 4 to 8 kHz | BKB phonetically balanced short sentences equalized to 65 dB SPL in terms of RMS | Between group comparison, repeated measures       | Dual-task paradigm (accuracy)           | There was a main effect of condition for both listening effort measured by accuracy [F(2, 68, 117.79) = 27.71, p < 0.001, η² = 0.39] and reaction time [F(2.28, 100.37) = 3.40, p < 0.05, η² = 0.07], suggesting that participants showed greater declines in secondary-task performance as spectral degradation increased. |
| Tun et al., 2009                                                     | 24 older adults (aged 67–80 years, mean: 73.9, SD: 4.1) divided into two groups (good hearing, poor hearing) | Semantically associated word lists and semantically unassociated word lists presented by female talker | Between group comparison                           | Dual-task paradigm (recall accuracy)    | Most important for our hypothesis, however, was the appearance of a significant Task × Hearing interaction, F[2, 88] = 6.25, p < 0.01, η² = 0.124, moderated by a significant Task × Age × Hearing interaction, F[2, 88] = 3.92, p < 0.05, MSE = 16.82, η² = 0.082. That is, hearing acuity had a significant negative impact on tracking accuracy with the effect primarily attributed to the older poor hearing group. |
| Ayasse et al., 2017                                                  | 20 older adults (age ranged 65 to 88 years, mean: 73.6) grouped by variable of age (10 normal hearing and 10 hearing-impaired) | An array of four pictures of objects displayed in the four corners of screen      | Between group comparison                           | Adjusted mean pupil size                | A one-way ANOVA yielded a significant effect of group on pupil diameter [F(2, 37) = 8.22, p = 0.001, η² = 0.308] which the hearing-impaired older adults showed a significantly greater increase in relative pupil size leading up to their eye fixation on the correct object picture as compared to both young and older adults with normal hearing (p = 0.003). |
| Rosemann & Thiel, 2020                                               | 19 adults with mild to moderate hearing loss (age mean: 64.63, SD: 6.3) | A matrix sentence test was used to identify the 80% speech-in-noise intelligibility of each group. | Between group comparison                           | CTMT flexibility                       | Cognitive flexibility significantly differed between groups: Hard of hearing group showed less flexibility in the trail making test (T(36) = 2.261; p = 0.03). |
| Bertoli & Bodmer, 2014                                               | 26 older adults with hearing loss (age ranged: 60–86 years, mean: 71.6) | Low-predictability sentences in a German version of the SPIN test presented in a constant background noise of 60 dB SPL | Between group comparison, repeated measures       | Correct final words                     | Two-way ANOVA with repeated measures revealed a significant main effect of correct final words [F(2, 84) = 22.31; p < 0.001; η² = 0.84]. |
|                                                                     |                                                   |                                                                                |                                                    | Novelty P3                              | Amplitudes increased steeply from the easy to the medium condition and decreased slightly towards the hard condition. Also, amplitudes of Cz differed significantly from Pz (p = 0.001) and Pz (p = 0.05). |
| Study | Participants | Test materials | Study design | Outcome measures | Main findings |
|-------|--------------|----------------|--------------|------------------|---------------|
| Giroud et al., 2017 | 13 adults with hearing aids (age ranged: 64–77 years, mean: 70.31, SD: 5.19) | Three non-sense syllables from the phoneme perception test | Between group comparison, repeated measures | P50 latency N1 latency P2 latency | For the P50 latency there was a main effect of group \( (F[2,36] = 4.37, p = 0.02, \eta^2_p = 0.20) \) revealing that across M2 and M3, the P50 latency was longer for the hearing-impaired compared to the group with normal hearing \( (p < 0.05) \). For the N1 latency, there was another main effect of group \( (F[2,36] = 9.59, p < 0.001, \eta^2_p = 0.35) \) similar to the P50 latency, showing that NHO had shorter latencies than both hearing impaired groups \( (p < 0.01) \). |
| Bernarding et al., 2013 | 24 middle-aged subjects with moderate hearing loss (age ranged: 43–57 years, mean: 51.12, SD: 5.53) | Consonant-vowel syllables (/ba/, /da/, /pa/, /de/, /bi/) were used as auditory stimuli. | Between group comparison, repeated measures | N1 latency | It is noticeable, that the WPSS is larger for the DSP than the ESP for all middle-aged subjects \( (p < 0.05, \text{one-way ANOVA}) \) in the time interval of interest corresponding to the N1 wave (approximately 70–160 ms). In the case of the young normal hearing subjects, the results of the WPSS revealed almost the same values. Note that a listening effort related comparison of both hearing impaired subject groups is not appropriate due to the individual loudness adjustment of the syllables for the moderate hearing-impaired subjects. Nevertheless, a significantly enhanced WPSS for the DSP is noticeable for the moderate hearing-impaired subjects. |
| Zekveld et al., 2009 | 30 middle-aged adults with hearing impairment (age ranged 46 to 69 years, mean: 57, SD: 7.0) and hearing aids \( (n = 8) \) either unilateral and/or bilateral | Continuous rating scale which translated to Dutch version (0 to 10 scale) | Between group comparison, repeated measures | Rating scale | The analysis indicated no group differences in the listening effort in the auditory tests \( (p > 0.05) \). In other words, processing the subtitles resulted in extra effort compared with auditory-only speech comprehension when the subtitles were presented delayed relative to the speech. |
| Rosemann & Thiel, 2019 | 19 participants (mean aged of 63.5 years, SD: 5.3) with symmetrical hearing loss | Listening effort questionnaire includes 17 questions about listening situations in everyday life from 0 (not effortful at all) to 10 (extremely effortful). | Between group comparison | Rating scale | Listening effort was significantly higher in the hearing-impaired group \( (3.27 \pm 1.65) \) than in the normal-hearing participants \( (2.26 \pm 1.19) \) \( (T(36) = 2.128; p = 0.02) \). High-frequency hearing loss and listening effort correlated significantly \( (r = 0.379; p = 0.019) \). |
| Mackesie et al., 2015 | 18 adults with sensorineural hearing loss (age ranged 22 to 79 years, mean: 58) | Respiration was measured using an elasticized respiration belt worn over the clothes at the level of the diaphragm. Electrocardiographic recordings to obtain HF-HRV were recorded using three electrodes attached to snap-on cables. Skin conductance was measured on the participant's non-dominant hand using two silver/silver chloride electrodes attached to. The electrodes were placed on opposite sides of the palm over the thenar and hypothenar muscles. | Between group comparison, repeated measures | Respiration rate HF-HRV Skin conductance reactivity | A repeated-measures ANOVA confirmed that there was no significant effect of hearing, relSNR, or interaction between these variables. HF-HRV decreased under the most difficult listening conditions for participants with hearing loss, but not for those with normal hearing. Mean HF-HRV for the hearing-loss group was significantly lower than that of those with normal hearing, but only at the two lowest relSNRs. After accounting for the effect of age, the main effect of hearing status remained significant \( (F(1,30) = 4.80, p = 0.03, \eta^2_p = 0.14) \). |
### Table 3. Summary of the extracted and synthesized data from 19 included articles (continued)

| Study                        | Participants                                                                 | Test materials                                                                 | Study design                              | Outcome measures | Main findings                                                                                                                                 |
|------------------------------|------------------------------------------------------------------------------|--------------------------------------------------------------------------------|-------------------------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| Alhanbali et al., 2019       | 116 participants age ranged 55 to 85 years (mean: 70, SD: 8) divided into   | Pupil size were measured using an Eye-Link 1000 with a sampling rate of 1,000 Hz. | Non-randomized controlled trials, Repeated measures. | Peak pupil size  | Pupil size increased significantly relative to baseline as participants attended to the speech, and reached a peak toward the end of the 3-sec speech stimulus. |
|                              | subgroups based on their hearing status: mild hearing-impaired (n = 42),     | Seven silver/silver chloride electrodes with a sintered surface were used. Three |                                                                                           | EEG              | There were increases in alpha activity toward the end of the retention period and an increase in alpha activity during speech presentation. Increased alpha activity associated with increased listening effort had mainly been observed over the parietal lobe. |
|                              | moderate hearing-impaired (n = 29), severe hearing-impaired (n = 8).          | positive electrodes were therefore placed over parietal scalp regions to capture |                                                                                           |                  |                                                                                                                                             |
|                              |                                                                              | task-related alpha activity: Pz, P3, and P4 based on the international 10 to 20 system. |                                                                                           |                  |                                                                                                                                             |
|                              |                                                                              | Two silver/silver chloride electrodes were attached to the index and the middle |                                                                                           | Skin conductance | The median score and IQR for the NASA Task Load Index were 34.16% (IQR: 26.25). For reaction time, the values were 1,945.86 ms (IQR: 540.71) and for skin conductance, 0.25 µs (IQR: 0.30). |
|                              |                                                                              | finger of the participant’s nondominant hand.                                   |                                                                                           |                  |                                                                                                                                             |
|                              |                                                                              | Monosyllabic digits (from 1 to 9) based on the Whispered Voice Test were used.  |                                                                                           |                  |                                                                                                                                             |
|                              |                                                                              | NASA Task Load Index which consists of six items including mental, physical,    |                                                                                           |                  |                                                                                                                                             |
|                              |                                                                              | temporal demand, perceived performance, effort, and frustration                 |                                                                                           |                  |                                                                                                                                             |
| Krueger et al., 2017         | 18 moderately hearing-impaired subjects (mean age: 74 years)                | Effort scale categorical units which rated between 1 (no effort) and 13 (extreme effort) | Between group comparison, repeated measures | Perceived        | The correlations between the SRT and the respective SNR value for the listening effort categories no effort to extreme effort were not significant for listeners with NH but were significant for the (unaided and aided) listeners with HI. |
|                              |                                                                              |                                                                                           |                                                                                           | listening effort |                                                                                                                                             |
| Holube et al., 2016          | 17 elderly hearing-impaired (age ranged from 52 to 85 years, mean: 73) with   | Subjective listening effort using a categorical rating scale with seven labeled | Between group comparison, repeated measures | Subjective       | For all four listening conditions, on average, 73% of the young normal-hearing subjects described associations of an imagined environment, compared with only 27% of the elderly hearing-impaired. These differences between young normal-hearing and elderly hearing-impaired listeners were statistically significant, \( \chi^2(1) = 21.71, p < 0.001 \). |
|                              | mild-to-moderate hearing loss                                                | categories and six intermediate steps between 1 (no effort) and 13 (extreme effort) |                                                                                           | listening effort |                                                                                                                                             |
| Study                  | Participants                                                                 | Test materials                                                                 | Study design                        | Outcome measures                                                                 | Main findings                                                                                                                                 |
|-----------------------|------------------------------------------------------------------------------|---------------------------------------------------------------------------------|-------------------------------------|----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| Picou & Ricketts, 2018| 18 adults (age ranged 50–80 years, mean: 66.6, SD: 8.5) with bilateral, symmetrical, sensorineural hearing loss | The CST lists which consisted of eight lists of 60 words with four talker babble noise | Non-randomized controlled trials, repeated measures | Dual-task paradigm (word recognition performance)                                | Analysis of word recognition performance revealed a significant main effect of hearing aid setting ($F_{[2, 13]} = 27.50, p < 0.001, \eta^2 = 0.81$). Also, analysis of response times revealed a significant main effect of hearing aid setting ($F_{[2, 13]} = 4.84, p < 0.05, \eta^2 = 0.43$). These data suggest that behavioral listening effort was improved with only the bilateral beamformer relative to the omnidirectional condition. |
| Huber et al., 2018    | 18 moderately hearing-impaired subjects (mean age: 74 years)                | The simplified version of MUSHRA scale was used. Listening effort scale were categorized into 7 categories from extremely stressful to effortless. | Between group comparison, repeated measures | Subjective listening effort                                                     | Listening effort ratings exhibit larger variations across subjects at 0 dB input SNR compared to ratings of other quality dimensions, especially for HI subjects. |
| Ricketts et al., 2019 | 32 adults aged 40–85 years (mean: 67.9 years) with mild sloping to moderate high frequency hearing loss | BKB Speech-in-Noise Test which presentation level of 83 dB SPL presented diotically to both ears through headphones. | Non-randomized controlled trials, repeated measures | Dual-task paradigm (response time)                                               | Word recognition was better with AV stimuli, $M = 16.76$ rau, 95% CI (14.17–19.36), and with two hearing aids, $M = 4.67$ rau, 95% CI (1.56–12.76). Analysis of response times during the secondary task revealed nonsignificant effects of modality, hearing aid fitting, and the Modality × Hearing Aid Fitting interaction. Analysis of subjective ratings revealed significant effects of modality, $F[1, 31] = 8.54, p < 0.01, \eta^2 = 0.22$, and hearing aid fitting, $F[1, 31] = 32.36, p < 0.0001, \eta^2 = 0.51$. |
| Alhanbali et al., 2018| 84 older adults with mild-to-severe hearing loss with and without hearing aid (i.e., 8 without hearing aid, 26 with unilateral hearing aid, and 50 with bilateral hearing aid) | Effort Assessment Scale questionnaire with 6 items                               | Non-randomized controlled trials, repeated measures | Subjective questionnaire (10-points Likert scale)                              | There was a significant positive correlation between worse speech recognition and greater effort and also greater fatigue ($R^2 = 0.56, p < 0.05$), that is, the need for a more positive SNR was associated with greater effort/fatigue. |

BKB: Bamford-Kowal-Bench, RMS: root-mean-square, DPRT: digital pursuit rotor tracking, YNH: young normal hearing, ONH: old normal hearing, OHI: old hearing-impaired, CTMT: Comprehensive Trail Making Test, SPIN: speech-in-noise test, SSN: speech-shaped noise, WPSS: wavelet phase synchronization stability, DSP: difficult syllabic paradigm, ESP: easy syllabic paradigm, HF-HRV: high-frequency heart-rate variability, EEG: electroencephalography, NASA: National Aeronautics and Space Administration, SRT: speech recognition threshold, SNR: signal-to-noise ratio, NH: normal-hearing, HI: hearing-impaired, CST: connected speech test, SSQ: Speech, Spatial and Qualities of Hearing Scale
2.490, $p = 0.001$) 가장 높은 효과 크기를 보였다. 자가 보고 측정과 주관적 측정에 따른 청취 노력의 결과 역시 각각 -1.280 (95% CI: -4.180~1.620, $p = 0.387$)와 1.221 (95% CI: 0.267~

2.175, $p = 0.012$)의 효과 크기를 보였으며 자가 보고 측정을 제외한 모든 측정 방법에서 통계적 유의미성이 검정되었다. 하위 군별 결과를 보다 자세히 확인하고자 결과 측정 방법에 따른

| Group by Study name | Subgroup within study | Std diff Standard in means | Std err | Variance | Lower limit | Upper limit | Z-Value | Value-Value |
|---------------------|-----------------------|-----------------------------|--------|----------|------------|------------|---------|-------------|
| objective Ayesa & Wingfield, 2018 objective | 4.569 | 0.718 | 0.510 | 3.616 | 3.970 | 3.636 | 0.000 |
| objective Tiu et al., 2009 (b) objective | 1.670 | 0.475 | 0.223 | 2.741 | 2.599 | 2.532 | 0.000 |
| objective Ayesa et al., 2017 objective | 4.773 | 0.728 | 0.530 | 3.347 | 6.200 | 6.558 | 0.000 |
| objective Giroud et al., 2017 (a) objective | 2.391 | 0.514 | 0.264 | 1.304 | 3.397 | 4.655 | 0.000 |
| objective Giroud et al., 2017 (b) objective | 2.617 | 0.534 | 0.280 | 1.569 | 3.664 | 4.897 | 0.000 |
| objective Giroud et al., 2017 (c) objective | 0.030 | 0.092 | 0.154 | -0.739 | 0.798 | 0.075 | 0.140 |
| objective Hernandiez et al., 2013 objective | 0.423 | 0.202 | 0.085 | -0.149 | 0.995 | 1.450 | 0.147 |
| objective Mackenzie et al., 2015 (a) objective | 1.577 | 0.400 | 0.160 | 0.793 | 2.360 | 3.043 | 0.000 |
| objective Mackenzie et al., 2015 (b) objective | 0.339 | 0.192 | 0.124 | -0.361 | 1.003 | 0.034 | 0.350 |
| objective Mackenzie et al., 2015 (c) objective | 2.632 | 0.472 | 0.227 | 1.608 | 3.566 | 5.522 | 0.000 |
| objective Albornoz et al., 2019 (a) objective | 2.577 | 0.307 | 0.094 | 1.976 | 3.178 | 8.398 | 0.000 |
| objective Albornoz et al., 2019 (b) objective | 1.960 | 0.263 | 0.060 | 1.444 | 2.477 | 7.441 | 0.000 |
| objective Albornoz et al., 2019 (c) objective | 1.064 | 0.214 | 0.046 | 0.674 | 1.515 | 5.102 | 0.000 |
| objective Bortol & Bodey, 2014 (b) objective | 0.652 | 0.439 | 0.192 | -0.208 | 1.518 | 1.487 | 0.137 |
| objective | 1.860 | 0.344 | 0.116 | 1.186 | 2.535 | 5.409 | 0.000 |
| self-report Danjardine & Dobhey, 2013 (a) self-report | -8.897 | 1.186 | 1.400 | -11.221 | -6.573 | -7.504 | 0.000 |
| self-report Rosenmann & Thiel, 2020 (a) self-report | 0.315 | 0.256 | 0.107 | -0.325 | 0.955 | 0.964 | 0.335 |
| self-report Zehrdid et al., 2009 self-report | 0.106 | 0.283 | 0.079 | -0.045 | 0.637 | 0.043 | 0.176 |
| self-report Rosenmann & Thiel, 2019 self-report | 0.702 | 0.334 | 0.112 | 0.041 | 1.357 | 2.100 | 0.000 |
| self-report Albornoz et al., 2019 (c) self-report | 2.976 | 0.357 | 0.114 | 2.315 | 3.637 | 8.823 | 0.000 |
| self-report Knepper et al., 2017 self-report | 1.815 | 0.780 | 0.108 | 0.827 | 3.343 | 2.328 | 0.020 |
| self-report Hofert et al., 2016 self-report | 2.762 | 0.853 | 0.272 | -1.901 | 7.438 | 3.240 | 0.000 |
| self-report Hofert et al., 2017 self-report | -5.987 | 1.094 | 1.197 | -11.011 | 2.033 | -5.446 | 0.000 |
| self-report Ricketts et al., 2019 (a) self-report | -0.473 | 0.382 | 0.146 | -0.231 | 0.275 | -1.239 | 0.215 |
| self-report Albornoz et al., 2018 self-report | 0.676 | 0.235 | 0.055 | 0.216 | 1.332 | 2.880 | 0.004 |
| self-report | -0.081 | 0.426 | 0.116 | -0.195 | 0.754 | -0.190 | 0.850 |
| subjective Danjardine & Dobhey, 2013 (b) subjective | 3.130 | 0.536 | 0.287 | 2.080 | 4.810 | 5.841 | 0.000 |
| subjective Wad et al., 2017 (a) subjective | 0.483 | 0.304 | 0.093 | 0.867 | 2.280 | 2.244 | 0.025 |
| subjective Wad et al., 2017 (b) subjective | 1.356 | 0.226 | 0.106 | 0.667 | 2.444 | 4.007 | 0.000 |
| subjective Tian et al., 2009 (a) subjective | 3.256 | 0.623 | 0.388 | 2.036 | 4.477 | 5.231 | 0.000 |
| subjective Rosenmann & Thiel, 2020 (b) subjective | 0.717 | 0.335 | 0.112 | 0.061 | 1.723 | 2.142 | 0.032 |
| subjective Bortol & Bodey, 2014 (a) subjective | 0.119 | 0.307 | 0.094 | -0.483 | 0.720 | 0.387 | 0.699 |
| subjective Albornoz et al., 2019 (d) subjective | 2.493 | 0.301 | 0.090 | 1.904 | 3.082 | 8.292 | 0.000 |
| subjective Pierre & Ricketts, 2016 (a) subjective | 1.994 | 0.389 | 0.041 | 0.609 | 3.288 | 2.888 | 0.030 |
| subjective Pierre & Ricketts, 2018 (b) subjective | 0.451 | 0.529 | 0.120 | -0.386 | 1.249 | 0.852 | 0.394 |
| subjective Ricketts et al., 2019 (a) subjective | 0.516 | 0.384 | 0.147 | -0.235 | 1.268 | 1.346 | 0.178 |
| subjective Ricketts et al., 2019 (b) subjective | 0.357 | 0.378 | 0.135 | -0.134 | 0.859 | 1.023 | 0.389 |
| subjective Ricketts et al., 2019 (c) subjective | 0.766 | 0.380 | 0.147 | 0.907 | 1.602 | 2.949 | 0.004 |
| Overall | 1.051 | 0.380 | 0.137 | -0.106 | 2.169 | 1.777 | 0.076 |

Figure 2. Forest plot (A) which showed effect size based on standardized difference in means, and funnel plot (B) which presented an asymmetrical shape of an all included 35 studies. CI: confidence interval.
효과를 추가적으로 분석하였다.

Figure 2B에 제시한 캐피카 그림은 좌측으로 채우치는 폭축 상을 보였으나, Egger의 regression asymmetry test 결과 축판
편향없이 없음을 확인하였다(intercept: 0.2986, standard error: 1.68084, p = 0.43160). 전체 문헌에 대한 이설상은 Cochran’s Q
test에서 유의미함을 보였으며(Q: 385.999, df: 34, p < 0.001). Higgins의 I² test 결과 역시 높은 수준의 이질성을 보였기 때문
에(I²: 91.91), 사전에 설정한 하위군 분석을 통해 청취 노력의
측정 방법(객관적 측정, 자기 보고 측정, 주관적 측정)에 따른
비교를 수행하였다.

### Figure 3. Results of subgroup analysis based on the measurement type of listening effort. Objective measurement (A), self-report measurement (B), subjective measurement (C). EEG: electroencephalography, CTMT: Comprehensive Trail-Making Test, CI: confidence interval.
Methods for Listening Effort Measurement

Effect of outcome measurement on listening effort

Figure 3A에 제시된 객관적 측정 방법에 따른 효과를 확인했을 때, 결과 측정 방법으로 동공 확장(pupil dilation)을 사용하는 것이 가장 큰 효과 크기에 동반되는 유의미성은 나타났다(effect size: 3.864, 95% CI: 2.284~5.443, p < 0.001). 자기 보고 측정의 결과, 측정 방법에 따른 효과는 Figure 3B에 제시되었다. 사용된 척도별 효과를 확인했을 때, 10점 척도를 사용하는 것이 1.02로 가장 큰 효과 크기를 보였으나, 동반적 유의미성은 확인되지 않았다(95% CI: -0.642~2.685, p = 0.229). 반면에, 10점 척도를 사용한 결과는 동반적 유의미성은 확인되었으나, 가장 낮은 효과 크기를 보였다(effect size: -8.897, 95% CI: -12.907~4.888, p < 0.001).

Figure 3C에 제시된 주관적 측정의 결과 측정 방법에 따른 효과는 단어 화상(word recall)을 확인한 정확도(effect size: 3.256, 95% CI: 0.710~5.802, p = 0.012)와 정확도(effect size: 2.163, 95% CI: 0.470~3.855, p = 0.012)를 사용하는 것이 큰 효과 크기를 보였다.

CONCLUSIONS

본 논문에서는 난청 성인 및 보청기 착용 성인의 청취 노력 측정에 사용되는 측정 방법의 정확성과 효율성을 확인하기 위해 체계적 문헌 고찰 및 메타 분석을 수행하였다. 난청 성인 및 보청기 착용 성인의 청취 노력에 대한 측정된 측면의 개별 문헌의 데이터를 추출 및 합성하며, 총 35건의 데이터를 측정 방법에 따라 세부 하위 그룹으로 분류한 후 메타 분석을 수행하였다. 객관적 측정과 주관적 측정은 난청 및 보청기 착용 성인의 청취 노력의 측정과에 있어서 효과가 확인되었으나, 이렇게 청취 노력의 측정과에 있어서 자기 보고 측정의 효과는 암시되지 않았다. Pichora-Fuller 등(2016)의 연구에서는 청취 노력의 주관적 측정 결과와 자기 보고 측정 결과의 불일치 상에 대해 보고한 바 있다. 즉, 청취 노력은 동기 부여, 인지적 자원, 노력의 3차원적인 인지 과정의 산출물이기 때문에 수많은 인지적 처리 과정의 영향을 받게 된다. 구체적으로 주관적 측정 결과와 자기 보고 결과의 불일치성은 개인의 강인한 상황에 영향을 받으며, 이것은 동기부여의 부정적인 영향을 미치게 된다. 동기 부여에 영향을 받는 높은 강인한 상황 및 경험에 대해 높은 동기부여를 기대할 수 있으며, 이는 높은 인지적 자원의 사용과 노력을 통해 높은 청취 노력의 발현을 보이게 되지만, 낮은 강인한 상황에서는 동일한 환경 및 경험에 대해 저조한 동기부여로 인해 인지적 자원의 사용이 낮아지게 되어 이는 높은 청취 노력에 악영향을 미친다.

청취 노력의 측정에 대한 가장 큰 효과를 보인 객관적 측정 방법에서는 동공 확장을 사용하여 청취 노력의 측정하는 것이 가장 효율적인 것으로 나타났으며, 주관적 측정 방법과 자기 보고 측정 방법에서는 각각 작업 기억을 활용한 화상 정확도와 10점 척도를 사용하는 것이 효율적인 것으로 확인되었다. 그러나 가장 큰 효과를 보였던 객관적 측정 방법만 아니라 주관적 측정과 자기 보고 측정 방법의 효과는 결과 측정 방법에 따라 청취 노력의 측정에 대한 효과에 상당한 차이를 보였다. 객관적, 주관적, 자기 보고 측정 방법에 따른 청취 노력의 결과 간 상관성을 비교 분석한 Alhanbali et al.(2019)의 연구에서는 청취 노력의 측정 방법과 환경에 따라 청취 노력이 다르게 나타날 수 있으며 이는 청취 노력이 다차원적 개념이기 때문에 발생 가능하다고 보고하였다. 이는 Kahneman(1973)의 인지 능력 모델과 Pichora-Fuller et al.(2016)의 FUEL에서의 보고와 동일하게 청취(listening)는 제시하는 소리를 청각 기관을 사용하여 들고 정통하여 분석하여 뇌로 전달하게 되며 뇌로 전달된 정보는 이후의 정의를 거쳐 기억(memory)의 형태로 저장되어 청각-인지적 상호작용의 과정으로 해석할 수 있다. 또한, 청취로 인해 유발되는 청각-인지적 상호작용을 중 인지적 기능의 전반에는 작업 기억, 집중력, 처리 속도 등이 포함되며 궤적적으로는 인지적 자원의 소모와 이로 인한 청취 노력의 야기시킨다(Alhanbali et al., 2019; Pichora-Fuller et al., 2016). 다시 정리해보면, 다차원적 개념의 청취 노력을 측정하기 위해 현재까지 동공 확장이나 EEG 등의 객관적 측정, 정확도와 반응 시간을 포함하는 주관적 측정, 대상자의 주관적인 응답에 의한 자기 보고 측정의 방법 등이 사용되고 있으나, 각 정상 방법의 개별적인 측정 결과는 청취 노력의 단편적인 정보만을 제공하기 때문에(Alhanbali et al., 2018; Alhanbali et al., 2019; Ayasse et al., 2017; Ayasse & Wingfield, 2018; Pichora-Fuller et al., 2016) 보다 종합적인 측정 과정에 이에 대한 분석이 필요하기로 사료된다.

임상적 관점에서 난청자 및 보청기 사용자의 청취 노력의 측정 방법에 대한 적용은 측정 방법에 따른 효과와 현실적인 환경에 따라 구분될 수 있다. 예를 들어, 가장 우수한 검사법으로 분석되는 동공 확장이나 EEG 등의 객관적 측정 방법은 특수 장비 및 시설을 구비한 상급병원 등의 임상 기관에서 적용이 가능하지만, 주관적 측정 방법인 작업 기억 검사 및 해석은 푸리야스의 전문가와 협력이 필요한 될 보다 효과적이고 효율적으로 난청자 및 보청기 사용자의 청취 노력을 정량적으로 측정할 수 있을 것으로 사료된다. 반면, 음성 검사를 활용하여 다양한 청취 환경에서 음성인지의 정확도 및 반응/응답 시간의 측정하는 주관적 측정 방법은 클리닉 및 전문 센터에서의 적용 가능하다. 또한, 난청자 및 보청기 사용자의 주관적으로 느끼는 자율을 정량적으로 평가하기 위한 자기 보고 측정 방법을 객관적,
주관적 측정법과 함께 사용하여 검사대상자의 중심의 주관적 결과를 보완하는 것 또한 매우 바람직하다.

결론적으로, 본 논문은 난청과 보청기 사용자에 요구되는 청취 노력의 측정 방법에 따른 효과를 체계적 문헌 고찰과 메타 분석에 근거하여 작성할 수 있다. 향후에는 높은 귀족 기반 연구(evidence-based research) 등 질적 우 수성이 확보된 연구를 통해 난청과 보청기 사용자의 청취 노력 측정에 대한 효율성 확인과 임상적 검증이 필요하다. 더불어 난청인 및 보청기 사용자 대상으로 어음인지도와 청취 노력 간 상관성을 정확하게 분석하고 이를 근거로 객관적 지표를 제시한다면 청취 노력의 과학적 측정 근거가 될 것이며 대상자의 어음인지도 평가, 보청기 적합, 인지적 자원의 사용에 대한 상담 등에 청취 노력의 임상적 활용도가 높을 것으로 생각된다.

중심 단어: 보청기, 난청, 청취 노력, 객관적 측정, 자가보고, 주관적 측정.

Ethical Statement
N/A

Acknowledgments
N/A

Declaration of Conflicting Interests
There are no conflict of interests.

Funding
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MST) (2019R1F1A1053060).

Author Contributions
Conceptualization: all authors. Formal analysis: Chanbeom Kwak. Funding acquisition: Woojae Han. Methodology: all authors. Supervision: Woojae Han. Validation: Woojae Han. Visualization: Chanbeom Kwak. Writing—original draft: Chanbeom Kwak. Writing—review & editing: Woojae Han. Approval of final manuscript: all authors.

ORCID iDs
Chanbeom Kwak https://orcid.org/0000-0001-5657-7536
Woojae Han https://orcid.org/0000-0003-1623-9676

REFERENCES
Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K. J. (2018). Hearing handicap and speech recognition correlate with self-reported listening effort and fatigue. *Ear and Hearing*, 39(3), 470-474.
Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear and Hearing*, 40(5), 1084-1097.
Ayasse, N. D., Lash, A., & Wingfield, A. (2017). Effort not speed characterizes comprehension of spoken sentences by older adults with mild hearing impairment. *Frontiers in Aging Neuroscience*, 8, 329.
Ayasse, N. D. & Wingfield, A. (2018). A tipping point in listening effort: Effects of linguistic complexity and age-related hearing loss on sentence comprehension. *Trends in Hearing*, 22, 2331216518790907.
Bernardini, C., Strauss, D. J., Hannemann, R., Seidler, H., & Corona-Stauss, F. I. (2013). Neural correlates of listening effort related factors: Influence of age and hearing impairment. *Brain Research Bulletin*, 91, 21-30.
Bertoli, S. & Bodmer, D. (2014). Novel sounds as a psychophysiological measure of listening effort in older listeners with and without hearing loss. *Clinical Neuroradiology*, 125(5), 1030-1041.
Centre for Reviews and Dissemination. (2009). *Systematic Reviews: CRD’s Guidance for Understanding Reviews in Health Care*. York University, Centre for Reviews and Dissemination. Retrieved from https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf.
Desjardins, I. L. & Doherty, K. A. (2013). Age-related changes in listening effort for various types of masker noises. *Ear and Hearing*, 34(3), 261-272.
Giroud, N., Lemke, U., Reich, P., Mattes, K. L., & Meyer, M. (2017). The impact of hearing aids and age-related hearing loss on auditory plasticity across three months - An electrocorticogram study. *Hearing Research*, 353, 162-175.
GRADE Working Group. (2004). Grading quality of evidence and strength of recommendations. *BMJ*, 328(7454), 1490.
Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557-560.
Holube, I., Haeder, K., Imbery, C., & Weber, R. (2016). Subjective listening effort and electrodermal activity in listening situations with reverberation and noise. *Trends in Hearing*, 20, 2331216516667734.
Huber, R., Bisitz, T., Gerkmann, T., Kiesling, J., Meister, H., & Kollmeier, B. (2018). Comparison of single-microphone noise reduction schemes: can hearing impaired listeners tell the difference? *International Journal of Audiology*, 57(sup3), S55-S61.
Kahneman, D. (1973). *Attention and Effort*. (1st ed.). Englewood Cliffs, NJ: Prentice-Hall.
Kang, H. (2015). Statistical considerations in meta-analysis. *Hanyang Medical Reviews*, 35(1), 23-32.
Krueger, M., Schulte, M., Zokoll, M. A., Wagener, K. C., Meis, M., Brand, T., et al. (2017). Relation between listening effort and speech intelligibility in noise. *American Journal of Audiology*, 26(3S), 378-392.
Mackersie, C. L., MacPhee, I. X., & Heldt, E. W. (2015). Effects of hearing loss on heart rate variability and skin conductance measured during sentence comprehension in noise. *Ear and Hearing*, 36(1), 145-154.
McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., et al. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International Journal of Audiology*, 53(7), 433-440.
Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1), 1.
Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, 37 Suppl 1, S55-S75.
Picou, E. M. & Ricketts, T. A. (2018). The relationship between speech recognition, behavioural listening effort, and subjective ratings. *International Journal of Audiology*, 57(6), 457-467.
PROSPERO. (2014). *International Prospective Register of Systematic Reviews*. York University, Centre for Reviews and Dissemination. Retrieved from http://cdn.elsevier.com/promis_misc/PROSPEROAnimal.pdf.
R Core Team. (2018). *R: A Language and Environment for Statistical Computing* [Software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: http://www.R-project.org/.
Ricketts, T. A., Picou, E. M., Shehorn, J., & Dittberner, A. B. (2019). Degree of hearing loss affects bilateral hearing aid benefits in ecologically relevant laboratory conditions. *Journal of Speech, Language, and Hearing Research*, 62(7), 2331216519791905.
Rosemann, S. & Thiel, C. M. (2019). The effect of age-related hearing loss and listening effort on resting state connectivity. *Scientific Reports, 9*(1), 3834-3850.

Rosemann, S. & Thiel, C. M. (2020). Neural signatures of working memory in age-related hearing loss. *Neuroscience, 429*, 134-142.

Tun, P. A., McCoy, S., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging, 24*(3), 761-766.

Ward, K. M., Shen, J., Souza, P. E., & Grieco-Calub, T. M. (2017). Age-related differences in listening effort during degraded speech recognition. *Ear and Hearing, 38*(1), 74-84.

Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., & Cox, L. C. (2006). Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *Journal of the American Academy of Audiology, 17*(7), 487-497.

Zekveld, A. A., Kramer, S. E., Kessens, J. M., Vlaming, M. S., & Houtgast, T. (2009). The influence of age, hearing, and working memory on the speech comprehension benefit derived from an automatic speech recognition system. *Ear and Hearing, 30*(2), 262-272.