

# Comparison of Dynamic Range of Clear Speech and Conversational Speech in Korean

Kyungju Lee<sup>1</sup>, Hongyeop Oh<sup>1</sup>, Suyeon Shin<sup>1</sup>, Sohee Heo<sup>1</sup>, In-Ki Jin<sup>2</sup>

<sup>1</sup>Department of Speech Pathology and Audiology, Graduate School, Hallym University, Chuncheon, Korea

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

## 한국어 기반 클리어스피치와 일반 발화에 따른 어음역동범위 비교

이경주<sup>1</sup> · 오홍엽<sup>1</sup> · 신수연<sup>1</sup> · 허소희<sup>1</sup> · 진인기<sup>2</sup>

한림대학교 일반대학원 언어병리청각학과<sup>1</sup>, 한림대학교 자연과학대학 언어청각학부 · 청각언어연구소<sup>2</sup>

**Purpose:** Speech dynamic range (DR) represents the range between maximum and minimum levels of speech across frequencies. The DR has been considered as an important factor in speech intelligibility. This study aimed to identify any differences in DR between Korean clear speech and conversational speech. **Methods:** A standardized sentence stimuli were recorded with manners of conversational and clear speech. The DR was quantified using cumulative histogram levels based on recorded voice data from 35 native Korean female participants with normal hearing. Then, the DRs were compared according to the production method after level normalization at 65 dB sound pressure level. **Results:** Maximum levels in the DR for Korean clear speech were greater than for Korean conversational speech in most frequency bands. Especially, differences of the DR between conversational speech and clear speech were between 0.47 dB (center frequency of 570 Hz) and 2.96 dB (center frequency of 1,170 Hz). **Conclusion:** The results of this study suggest that the DR for Korean clear speech has acoustical characteristics that are favorable to speech intelligibility, compared to the DR for conversational speech.

**Key Words:** Clear speech, Speech intelligibility, Dynamic range of speech.

**Received:** January 10, 2018 / **Revised:** February 12, 2018 / **Accepted:** March 10, 2018

**Correspondence:** In-Ki Jin, Division of Speech Pathology and Audiology, Research Institute of Audiology and Speech Pathology, College of Natural Sciences, Hallym University, 1 Hallimdaehak-gil, Chuncheon 24252, Korea

**Tel:** +82-33-248-2221 / **Fax:** +82-33-256-3420 / **E-mail:** inkijin@hallym.ac.kr

## INTRODUCTION

클리어스피치(clear speech)는 소음하 상황 등 어려운 청취 상황에서 효과적인 의사소통을 위해 사용할 수 있는 의사소통 전략 중 하나로 알려져 있으며(Picheny et al., 1985), 음의 높낮이 범위(pitch) 확장, 핵심 어휘의 강도(intensity) 증가, 발화 속도(speech rate)의 감소를 통한 음향학적 조음 수정 전략을 의미한다(Smiljanić & Bradlow, 2007). 몇몇 연구들에서 클리어스피치는 일상생활에서 사용하는 일반 발성법에 비해 향상된 어음인지도(speech intelligibility) 수행력을 보인다고 보고하고 있다(Picheny et al., 1985; Smiljanić & Bradlow, 2007). 예를 들어, 영어를 모국어로 사용하는 난청인의 경우, 클리어스피치의 사용은 일반 발성법의 사용 시와 비교했을 때 17% 정도 높은

어음인지도 수행력을 보이는 것으로 나타났다(Picheny et al., 1985). 클리어스피치는 일반 발성법과 비교했을 때 상대적으로 느린 발화 속도로 발생되어 분절음의 발생 정확도가 증가되고 이러한 음향학적 변화는 어음인지도의 향상에 기여하는 것으로 보고되고 있다(Picheny et al., 1985). Krause & Braida (2002)는 보통 발화 속도와 느린 발화 속도에서 클리어스피치와 일반 발화 간 어음인지도를 측정하였는데 두 가지 발화 유형 모두에서 발화 속도가 느릴 때 5% 정도 더 높은 어음인지도를 보인다고 보고하였다. 또한 클리어스피치는 일반 발성법과 비교해 보았을 때, 중주파수 대역인 1,000~3,000 Hz 사이에서 장기평균어음스펙트럼(long-term averaged speech spectra)의 에너지 증가가 두드러지는 것으로 보고하고 있다(Krause & Braida, 2004). 이러한 중주파수 대역(1,000~3,000 Hz)의 에너지 증가

로 인한 어음인지도의 향상 효과는 약 6% 정도인 것으로 나타났다(Krause & Braida, 2009). 이외에도 클리어스피치는 모음의 확장성(vowel expansion)이 두드러지며(Ferguson & Kewleyport, 2002), 시간축포락선(temporal envelope)이 변조되는 등(Krause & Braida, 2004)의 음향학적 특성이 나타나는 것으로 알려져 있다.

대부분의 클리어스피치 연구는 영어를 기반으로 이루어졌지만, 몇몇 연구들은 크로아티아어, 핀란드어, 한국어와 같은 다양한 언어를 기반으로 클리어스피치의 음향학적 특성을 분석하였다(Cho et al., 2011; Granlund et al., 2012; Smiljanić & Bradlow, 2005). 예를 들어, 한국어를 대상으로 한 클리어스피치 연구에서는 느린 발화 속도와 모음의 확장성과 같은 음향학적 특성이 영어 클리어스피치와 유사하게 나타난다고 보고하였다(Cho et al., 2011). 또한 크로아티아어 클리어스피치도 일반 발성법에 비해 느린 말속도와 모음의 확장성 등이 두드러진다고 보고하고 있다(Smiljanić & Bradlow, 2005). 이처럼 한국어를 비롯한 여러 언어들에서 클리어스피치의 몇몇 음향학적 특징들이 보고되고 있지만 영어 클리어스피치에서 분석된 특징들에 비해서는 극히 일부 특징들만이 분석되었다.

본 연구에서는 보편적인 음향 특징에 해당하는 어음역동범위(dynamic range)의 분석을 통해 한국어 클리어스피치의 특징을 분석해 보고자 하였다. 어음역동범위는 주파수별 어음 에너지 분포 중 최소어음레벨(minimum level)과 최대어음레벨(maximum level)의 사이를 나타내며 어음인지에서 중요한 역할을 한다. 다시 말해 어음역동범위에서 청취자의 청력 역치 이상의 부분만이 어음인지도에 기여할 수 있다. 이렇듯 어음역동범위는 어음인지도 수행력을 결정하는 데 중요한 역할을 하기 때문에 본 연구를 통하여 한국어 클리어스피치의 어음역동범위 특성을 살펴보고자 한다.

본 연구의 목적은 한국어 클리어스피치와 일반 발성법 간에 어음역동범위를 측정하여 발성법에 따른 어음역동범위의 차이를 비교하고자 하였다. 본 연구의 결과는 한국어 클리어스피치의 음향학적 특징이 어음인지도의 향상에 기여할 수 있는지에 대한 기초 정보를 제공할 것이다.

## MATERIALS AND METHODS

### Participants

본 연구는 한국어를 모국어로 사용하는 35명의 여성 화자(평균 연령: 21.5세, 연령 분포: 20~28세)를 대상으로 진행하였다. 모든 대상자의 청력 역치는 250~8,000 Hz의 주파수 대역에서 20 dB HL 이하로 정상 범위의 청력 역치를 가지고 있었다. 본 연구는 녹음을 통한 목소리의 분석이 주요한 방법이기

때문에 조음과 발성에 문제가 없는 대상으로 한정하여 피검자를 모집하였다. 실험의 진행에 앞서 모든 연구에 대한 절차 및 내용은 임상실험윤리위원회(Institute Review Board, IRB)의 승인(HIRB-2016-001)을 받았으며 모든 연구가 완료된 후 참여자들에게 연구에 대한 수고비를 지급하였다.

### Recording and analysis procedure

본 연구에서 적용한 한국어 클리어스피치의 훈련 방법은 선행 논문에서 보고하고 있는 영어 클리어스피치 훈련 방법과 유사하게 진행하였다(Picheny et al., 1985; Smiljanić & Bradlow, 2005). 녹음 문장은 한국어 문장검사목록(Korean Speech Perception In Noise)의 표준화된 240개 문장을 사용하였으며(Kim et al., 2000), 모든 대상자는 클리어스피치와 일반 발성법의 두 가지 방식으로 전체 문장을 녹음하였다. 클리어스피치는 일반 발화법에 비해 느린 말 속도를 가지고 핵심 단어와 절 사이에 휴지기(pause)를 가진다(Picheny et al., 1986). 또한 최대한 정확하게 발성하며 생동감 있는 목소리(lively)를 유지하며 말끝을 흐리지 않는다(Caissie et al., 2005; Kricos, 2006; Moon & Lindblom, 1994; Picheny et al., 1986). 이러한 점들을 잘 반영할 수 있도록 클리어스피치 녹음 이전에 20~40분 사이의 연습 시간을 가졌다. 클리어스피치를 녹음할 때는 배경 소음이 있는 상황에서 외국인과 이야기하듯이 읽도록 안내하였으며, 일반 발성법의 녹음은 평상시에 말하는 것처럼 익숙한 사람들과 대화하듯이 읽도록 안내하였다. 음성 샘플의 녹음은 이중벽 방음실(double-wall sound booth) 내에 설치된 음성분석프로그램(Computerized Speech Lab; PENTAX Medical, Montvale, NJ, USA)과 마이크(Sennheiser e-835s; Sennheiser, Wedemark, Germany)를 통해 진행하였다. 표본화 주파수는 44,100 Hz로 설정하였고 양자화는 16 비트(bit)로 설정하여 녹음을 진행하였다. 문장을 녹음하기 전에, 대상자들의 긴장을 풀어주기 위해 몇 가지 문장을 연습으로 읽게 하였다. 대상자는 의자에 편안하게 앉은 상태에서 입은 마이크로부터 10 cm 거리를 유지하여 녹음을 진행하였다. 연구 참여자 중 18명은 한국어 문장검사목록을 일반 발성법으로 녹음을 한 이후에 클리어스피치 방식으로 녹음을 진행하였으며, 나머지는 클리어스피치 방식으로 녹음을 진행한 후 일반 발성법의 녹음을 진행하였다. 모든 참여자들은 녹음 도중 휴식을 원할 경우 수시로 휴식을 제공하였다. 연구 참여자당 총 녹음 시간은 90분 이내였다.

녹음된 클리어스피치와 일반 발성법의 어음역동범위를 산출하기 위하여 Adobe Audition CS 6 (Adobe Systems Inc., San Jose, CA, USA)을 이용하여 음성 파일들의 휴지기를 제거하였다. 목음이 제거된 녹음 파일들을 Matlab (version R2017a; MathWorks Inc., Natick, MA, USA)을 통해 누적 히스토그램

레벨(cumulative histogram levels)을 산출하였다(Jin et al., 2014; Kates, 2013). 누적 히스토그램 레벨 도출은 아래의 절차대로 진행하였다. 첫째, 녹음 데이터의 평균 실효값(root-mean square)을 65 dB sound pressure level (SPL)로 정규화(normalization)하였다. 둘째, 150~8,600 Hz의 중심주파수 범위를 갖는 21개의 임계 대역(critical bands)에서 신호 포락선(signal envelope)을 계산하였다. 이러한 21개의 주파수 대역 범위는 어음인지지수(Speech Intelligibility Index)를 계산할 때의 임계 대역 범위를 기준으로 하였다(American National Standards Institute, 1997/R2012). 어음인지지수는 대상자의 가청 영역을 주파수 대역별로 산출화하여 어음인지를 예측하는 모델이다(American National Standards Institute, 1997/R2012). 셋째, 매트랩을 이용하여 1000분의 1초 단위로 계산되는 상승 여현창(raised cosine window)을 통해 신호 포락선을 부드럽게 곡선화하였다. 넷째, 곡선화된 신호 포락선을 창 길이(window length)가 50% 중첩되게 하여 부표본화하고 dB SPL 단위로 변환시켰다. 다섯째, 누적 신호 포락선 분포 레벨을 데시벨 신호 포락선 히스토그램으로 계산하였다. 마지막 단계로, 각 주파수 대역에서의 최종 어음역동범위는 각 연구 참여자들의 누적 히스토그램 값을 평균하여 산출하였다. 누적 히스토그램 레벨은 주파수 대역 내에서 어음의 누적 신호 포락선의 분포를 보여준다. 예를 들어, 99% 누적히스토그램 레벨(L99)은 측정된 어음 신호에서 최고 강도 레벨(peak level)을 포함하거나 그 이하의 99% 부분을 나타내며, 30% 누적히스토그램 레벨(L30)은 측정된 어음 신호에서 최소 강도 레벨을 포함하거나 그 이상의 30% 부분을 나

타낸다. 본 연구에서 설정한 어음역동범위는 최대 레벨에서 최소 레벨을 뺀 값을 나타내며(L99-L30), 이 방식은 국제어음시험신호(International Speech Test Signal)에서 효과적인 어음역동범위로 제시된 기준이다(Holube et al., 2010).

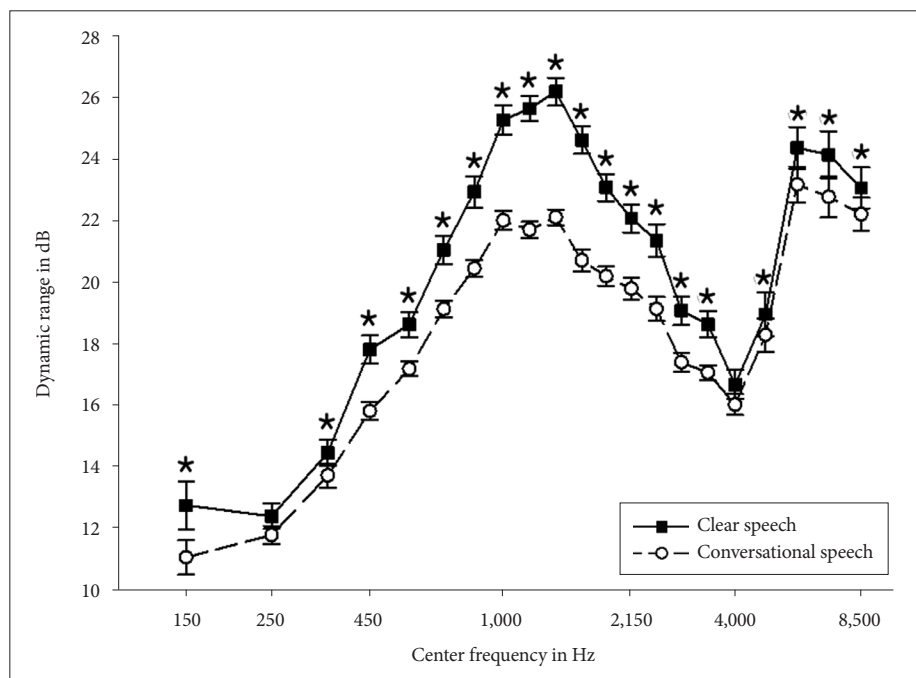
**Statistical analysis**

본 연구는 녹음 데이터를 가지고 두 가지 발성법 종류(클리어스피치, 일반 발성)에 따른 어음역동범위, 최대어음레벨, 최소어음레벨의 유의미한 차이 여부를 확인하였다. 통계분석은 SPSS-22 (IBM Corp., Armonk, NY, USA) 소프트웨어를 통해 분석하였고 대응표본 *t* 검정(paired sample *t*-test)을 이용하여 분석을 실시하였다.

**RESULTS**

발성법 종류(클리어스피치, 일반 발성)에 따른 어음역동범위는 Figure 1에 제시하였다. 클리어스피치의 어음역동범위는 19개의 주파수 대역에서 일반 발성법의 어음역동범위보다 유의미하게 넓은 것으로 분석되었다( $p < 0.05$ ). 통계적으로 유의미한 차이가 나는 주파수 대역에서 클리어스피치와 일반 발성법 간의 어음역동범위 차이는 최소 0.67 dB(중심주파수: 4,800 Hz)에서 최대 4.09 dB(중심주파수: 1,370 Hz) 사이였다.

Figure 2에서는 발성법 종류(클리어스피치, 일반 발성)의 최대어음레벨(L99)과 최소어음레벨(L30)을 나누어 비교하였다. 최대어음레벨(L99)의 경우, 두 개의 주파수 대역(중심주파수:



**Figure 1.** The DR for clear (square) and conversational (circle) speech as a function of 21 band frequencies. Asterisks indicate frequency bands wherein the DR of clear speech differed significantly from that of conversational speech. DR: dynamic range.

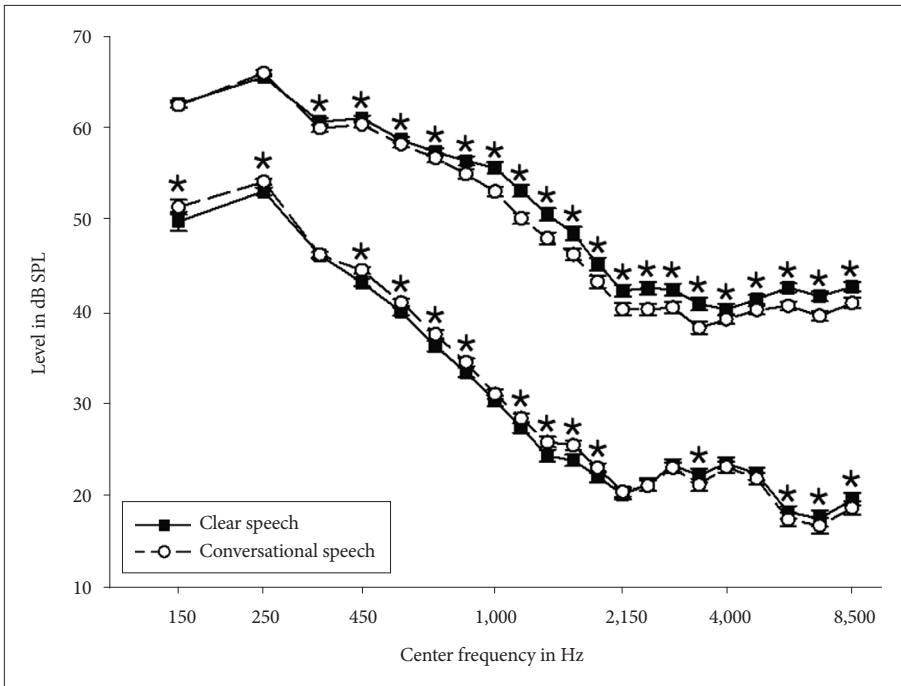


Figure 2. Maximum and minimum levels of dynamic range for clear (square) and conversational (circle) speech as a function of 21 band frequencies. Asterisks indicate frequency bands wherein the maximum (upper) and minimum (lower) levels of clear speech differed significantly from that of conversational speech. SPL: sound pressure level.

150, 250 Hz)을 제외하고 19개의 주파수 대역에서 클리어스피치가 일반 발성법보다 더 높은 최대어음레벨(L99)을 보였으며 ( $p < 0.05$ ), 그 차이는 최소 0.47 dB(중심주파수: 570 Hz)에서 최대 2.96 dB(중심주파수: 1,170 Hz) 사이였다. 최소어음레벨(L30)의 경우, 10개의 중심주파수 대역(중심주파수: 150, 250, 450, 570, 700, 840, 1,170, 1,370, 1,600, 1,850 Hz)에서 클리어스피치가 일반 발성법보다 더 낮은 최소어음레벨(L30)을 보였다( $p < 0.05$ ). 이와 반대로 클리어스피치의 최소어음레벨(L30)은 4개의 중심주파수 대역에서 클리어스피치가 일반 발성법보다 더 높은 최소어음레벨(L30)을 보였다( $p < 0.05$ ). 이렇듯 통계적으로 유의미한 주파수 영역에서 클리어스피치와 일반 발성법의 최소어음레벨(L30)의 차이는 최소 0.76 dB(중심주파수: 7,000 Hz)에서 최대 1.63 dB(중심주파수: 1,600 Hz) 사이였다.

DISCUSSIONS

본 연구에서는 한국어 클리어스피치가 일반 발성법보다 더 넓은 어음역동범위를 가지고 있음을 확인하였다. 본 연구의 결과를 어음역동범위가 어음인지수에서의 예측에 미치는 영향과 관련 지어 연구한 선행 논문의 결과에 적용하였을 때, 한국어 클리어스피치가 일반 발성법에 비해 어음인지도에 유리한 음향학적 특성을 가지고 있음을 예측할 수 있다. Jin et al.(2017) 연구에서는 어음역동범위가 10 dB에서 60 dB로 넓어지면 동일한 가청 영역에서 어음인지수에서의 어음인지도 수행력 점수도 높아진다고 보고하였다. 어음역동범위가 넓어지면 더 높은 최

대어음레벨(L99)을 가진 평균 실효값 레벨을 포함하기 때문에 신호대잡음비(signal-to-noise ratio)가 높아지게 되고 어음인지수에서의 전이함수(transfer function)에서 높은 어음인지도 점수를 산출한다. 예를 들어, 현재 어음인지수에서의 효과적인 어음역동범위는 30 dB로 평균 실효값 레벨을 기준으로  $\pm 15$  dB을 나타내는데(American National Standards Institute, 1997/R2012), 만약 어음역동범위가 평균 실효값 레벨을 기준으로 -15 dB 에서 +25 dB을 나타내는 40 dB로 넓어지게 된다면 더 높은 평균 실효값 레벨을 포함하기 때문에 신호대잡음비가 높아지게 될 것이다. 또한 Jin et al.(2017) 연구에서는 한국어의 절대 전이함수(absolute transfer function) 그래프에서 0.52의 어음인지수 값을 기준으로 어음역동범위가 30 dB에서 40 dB로 넓어질 때, 어음인지도 수행력 점수가 대략 27%에서 57.5%로 증가한다고 보고하였다. 이러한 예시를 통해 어음역동범위 값이 1 dB 증가할 때, 어음인지도 수행력은 평균적으로 3.05% 증가하는 것을 확인할 수 있다.

클리어스피치의 어음인지도 향상 효과는 많은 선행 논문들에서 보고하고 있다. Bradlow & Bent(2002) 연구에서는 32명의 비원어민과 32명의 원어민을 대상으로 다양한 신호대잡음비 상황에서 클리어스피치와 일반 발화 간의 어음인지도 효과를 비교하였다. 그 결과, 비원어민과 원어민 집단 모두에서 클리어스피치를 사용했을 때 일반 발화법보다 더 높은 어음인지도 수행력을 보이는 것으로 나타났다. 또한 Bradlow et al.(2003) 연구에서는 건청아동 36명과 학습장애를 가진 아동 63명을 대상으로 클리어스피치와 일반 발화 간의 어음인지도 효과를 비

교하였다. 그 결과 클리어스피치 상황에서 건청아동 그룹에서는 8.8%, 학습장애아동 그룹에서는 9.2% 정도 더 높은 어음인지도 수행력을 보이는 것으로 나타났다. 선행연구들의 결과들은 영어 클리어스피치가 어음인지도 향상에 기여한다는 것을 증명하고 있다. 본 연구의 결과는 한국어 클리어스피치도 어음인지도 향상에 기여할 수 있다는 가능성을 보여주는 만큼 한국어의 음향적 특성을 고려한 클리어스피치 연구를 추가적으로 진행한다면 한국인을 위한 효과적인 의사소통 전략으로 적용될 수 있을 것이다.

본 연구의 결과는 클리어스피치의 어음역동범위가 전주파수 대역에서 일반 발성법의 어음역동범위보다 더 넓은 분포를 보였고, 이는 대부분 클리어스피치의 최대어음레벨(L99) 값이 일반 발성법의 최대어음레벨(L99) 값보다 높다는 점(최소 0.47 dB~최대 2.96 dB)에서 한국어 클리어스피치는 일반 발성법에 비하여 어음인지도에 유리한 음향학적 특성을 가지고 있음을 시사한다.

**중심 단어** : 클리어스피치 · 어음인지 · 어음역동범위.

### Acknowledgments

이 성과는 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2015R1C1A1A01052458).

### REFERENCES

- American National Standards Institute (ANSI). *Methods for Calculation of the Speech Intelligibility Index*. ANSI S3.5-1997 (R2012). New York, NY: ANSI.
- Bradlow, A. R. & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, 112(1), 272-284.
- Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research*, 46(1), 80-97.
- Caissie, R., Campbell, M. M., Frenette, W. L., Scott, L., Howell, I., & Roy, A. (2005). Clear speech for adults with a hearing loss: Does intervention with communication partners make a difference? *Journal of the American Academy of Audiology*, 16(3), 157-171.
- Cho, T. H., Lee, Y. J., & Kim, S. H. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, 39(3), 344-361.
- Ferguson, S. H. & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 112(1), 259-271.
- Granlund, S., Hazan, V., & Baker, R. (2012). An acoustic-phonetic comparison of the clear speaking styles of Finnish-English late bilinguals. *Journal of Phonetics*, 40(3), 509-520.
- 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.
- Jin, I. K., Kates, J. M., & Arehart, K. H. (2014). Dynamic range for speech materials in Korean, English, and Mandarin: A cross-language comparison. *Journal of Speech, Language, and Hearing Research*, 57(5), 2024-2030.
- Jin, I. K., Kates, J. M., & Arehart, K. H. (2017). Sensitivity of the Speech Intelligibility Index to the assumed dynamic range. *Journal of Speech, Language, and Hearing Research*, 60(6), 1674-1680.
- Kates, J. M. (2013). Improved estimation of frequency importance functions. *The Journal of the Acoustical Society of America*, 134(5), EL459-EL464.
- Kim, J. S., Pae, S. Y., & Lee, J. H. (2000). Development of a test of Korean speech intelligibility in noise (KSPIN) using sentence materials with controlled word predictability. *Speech Sciences*, 7(2), 37-50.
- Krause, J. C. & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *The Journal of the Acoustical Society of America*, 112(5 Pt 1), 2165-2172.
- Krause, J. C. & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, 115(1), 362-378.
- Krause, J. C. & Braida, L. D. (2009). Evaluating the role of spectral and envelope characteristics in the intelligibility advantage of clear speech. *The Journal of the Acoustical Society of America*, 125(5), 3346-3357.
- Kricos, P. B. (2006). Audiologic management of older adults with hearing loss and compromised cognitive/psychoacoustic auditory processing capabilities. *Trends in Amplification*, 10(1), 1-28.
- Moon, S. J. & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America*, 96(1), 40-55.
- Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, 28(1), 96-103.
- Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, 29(4), 434-446.
- Smiljanić, R. & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, 118(3 Pt 1), 1677-1688.
- Smiljanić, R. & Bradlow, A. R. (2007). Proceedings of the 16th International Congress of Phonetic Sciences: *Clear Speech Intelligibility: Listener and Talker Effects*. Saarbrücken, Saarland: Saarland University.