INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Speaking with a KN95 Face Mask: ASR Performance and Speaker Compensation

Sarah E. Gutz[1], *Hannah P. Rowe*[2], *Jordan R. Green*[1,2]

[1]Harvard University, USA
[2]MGH Institute of Health Professions, USA
gutz@g.harvard.edu, hrowe@mghihp.edu, jgreen2@mghihp.edu

Abstract

The increasing prevalence of face masks in the United States due to the COVID-19 pandemic necessitates serious consideration of the functional impact of wearing a mask on speech. This study considers how the presence of a KN95 mask affects the performance of a commercial ASR system, Google Cloud Speech. We present evidence that wearing a mask does not impact ASR performance at the sentence level. Moreover, speakers may be naturally adapting to the mask by increasing their vowel space area. However, when speakers intentionally altered their speech by speaking clearly or loudly (though not slowly), ASR performance improved. These findings suggest that ASR users can employ speech strategies to achieve better ASR results when wearing a mask. Beyond healthy speakers, our study has implications for mask-wearing ASR users with otherwise reduced speech intelligibility.

Index Terms: automatic speech recognition, medical mask, speech strategies, speech production

1. Introduction

Surgical masks and other face coverings have become increasingly prevalent in the general population since the start of the COVID-19 pandemic. Even prior to the pandemic, the Centers for Disease and Control Prevention (CDC) and Occupational Safety and Health Administration (OSHA) recommended masks in medical settings and construction sites, as face coverings have been shown to filter particles and prevent the spread of infection [1]. Now, however, in light of COVID-19, mask-wearing has become a regular way of life [2], [3]. While masks are crucial for containing potentially infectious aerosols during speech [4]–[6], they may deleteriously impact automatic speech recognition (ASR) performance. Reduced ASR performance could negatively affect individuals who use ASR to control electronic devices, interact with their environment due to mobility issues, or dictate notes, as in a medical setting. However, the exact extent to which mask-wearing impacts ASR performance is currently unknown. To maximize ASR performance for speakers wearing masks, there is also a need to understand explicit compensatory strategies that could help users optimize ASR accuracy.

Acoustically, masks have been found to disproportionately impact higher frequencies [7]–[10] and decrease intensity overall [8], [11]. Goldin et al. (2020) found that frequencies in the 2-7kHz range were attenuated by 3-4 dB for simple medical masks and close to 12dB for N95 masks [10]. The authors of this study also played prerecorded speech through surgical masks and N95 masks and found that the masks acted as low-pass filters. Because the researchers tested the acoustics of masks on prerecorded speech, they could remove the impact of possible compensatory strategies, such as increased loudness [12]. Also using prerecorded speech, Bottalico et al. (2020) found that fabric masks attenuated signal intensity by 4.2 dB, surgical masks by 2.3 dB, and N95 masks by 2.9 dB [13]. While these studies do not demonstrate the functional impact of masks, they elucidate the pure acoustic effect of these masks without the influence of speakers' compensating for the presence of the mask (e.g., speaking loudly to overcome the filter effect). Despite the acoustic effects of mask-wearing and evidence that machine learning systems can detect these changes [8], [14], [15], the precise impact of wearing a mask on ASR has not been fully established.

Given the association between perceptual intelligibility and ASR performance [16], research regarding the impacts of wearing a mask on intelligibility could inform our understanding of their impacts on ASR. Indeed, previous work has shown reductions in speech intelligibility due to alterations in the acoustic signal and, to a large extent, removal of visual information [11], [17]–[19]. Palmiero et al. (2016) noted an impact of prerecorded masked speech on perceptual intelligibility when listeners made their judgements using the audio file only [17]. Additionally, the authors demonstrated that medical masks decreased speech intelligibility by 3-17%. Radonovich et al. (2010) similarly found that wearing different masks impacted speech intelligibility by 1-17% [18]. Bottalico et al. (2020) report that cloth, surgical, and N95 masks reduced intelligibility of prerecorded speech by 12-16% [13].

Although recent work has reported on the impact of masks on intelligibility, many studies have filtered prerecorded speech through the mask, thereby removing the impact of speaker compensation on outcome measures. As Asadi et al. (2020) found, people may naturally alter how they speak while wearing a mask by speaking more loudly [12]. Such natural compensation could influence speech intelligibility beyond what we would expect from applying an acoustic filter alone to speech. Research has shown that strategies such as Clear, Loud, and Slow speech can significantly increase speech clarity and intelligibility [20]–[24]. Given the potential for automatic compensation and its impact on intelligibility, there is a critical need to examine the effects of mask-wearing on ASR performance in conjunction with clinical speech strategies.

We sought to address this gap by recording neurologically healthy speakers while wearing a mask and not wearing a mask. Additionally, to judge the benefits of compensation on masked speech, we recorded speakers using the three different speaking strategies with a mask. Determining the impact of compensatory strategies used by mask-wearers on ASR performance could a) further our understanding of natural compensatory strategies speakers employ while wearing a mask and b) provide guidance on how to improve ASR performance.

http://dx.doi.org/10.21437/Interspeech.2021-99

2. Methods

2.1. Participants

Participants were 19 healthy adults (14 females), ages 20-36 (median 26 years), who spoke North American English natively. Participants had no history of neurological disease and had hearing and vision skills adequate for completing the tasks.

2.2. Data collection

2.2.1. *Conditions*

This study examined the effects of three commonly used clinical speech strategies: Clear Speech, Loud Speech, and Slow Speech. All three speech strategies have been shown to improve speech intelligibility [20]–[24]. Participants completed speaking tasks in five total conditions: Mask, No Mask, Clear, Loud, and Slow. During the mask, Clear, Loud, and Slow conditions, participants wore a KN95 mask provided by the experimenter. KN95 masks are better regulated than fabric masks, yet more available to members of the public than N95 masks. Additionally, pilot studies our lab conducted show that KN95 masks have an acoustic impact on speech closer to that of fabric and N95 masks than surgical masks. To capture their habitual speech, participants completed the Mask and No Mask conditions first, with the order counterbalanced across participants. Participants then completed the three speaking strategy conditions also in a counterbalanced order. Speakers were given instructions identical to those in [25], [26].

2.2.2. *Audio recording and sound pressure level*

Audio (16-bit, 50 kHz mono) was recorded using a head-fixed microphone, approximately two inches from the speaker's mouth. Audio was downsampled to 48 kHz sampling rate for use with the ASR system. Sound pressure level (SPL) expressed in (dB) was recorded with A-weighting using an SPL meter fixed two feet away from the speaker and calibrated using two semitones.

2.2.3. *Speech stimuli*

In each condition, for measures of ASR recognition, speakers read 11 sentences of length five to 15 words from the Sentence Intelligibility Test (SIT) [27]. To calculate vowel formants and vowel space measurements, speakers read a story containing corner vowels /i, u, ae, a/ in /b-t/ context within sentences [28]. For each condition, there were three, four, five, and 10 tokens of the vowels /i, u, ae, a/, respectively. Speakers also read the Bamboo Passage in each condition, a paragraph that has been validated for measures of speaking rate and allows for intensity calculations on the same stimulus across conditions [29].

2.3. Outcome measures

2.3.1. *Automatic speech recognition (ASR)*

Google Cloud Speech was used to transcribe each recorded SIT sentence. Word Error Rate (WER) was calculated as minimum-edit distance between the ASR transcript and the ground-truth target SIT sentence and presented as a percentage of the ground-truth sentence that was mistranscribed. Speech was transcribed both as recorded and after being mixed with multi-talker babble. Noise mixing has been found to attenuate ceiling effects of intelligible speakers for human transcription intelligibility [20], [30]. This mixing allowed us to detect subtle differences in speaker intelligibility that may be minimal in ideal recording conditions but substantial in more ecological or noisy conditions. The noise added for each speaker was mixed at a signal-to-noise (SNR) of -1 dB relative to the No Mask SIT production. Every speech sample for a given speaker was mixed with the same absolute level of noise so that we could compare the functional impact of the different conditions in a consistent environment. A pilot study in our lab showed that a -1 dB SNR minimized both ceiling and floor effects of ASR transcriptions.

2.3.2. *Vowel space area*

F1 and F2 were evaluated as mean measurements of the middle 30 ms of each corner vowel (i.e., /i, ae, u, a/) produced. Vowel space area (VSA) was calculated as the mean quadrangle area using the phonR package [31].

2.4. Statistical analyses

To evaluate the effect of the conditions on outcome measures, we used mixed-effects models with condition (Mask, No Mask, Clear, Loud, Slow) as the predictor and participant as a random effect. *P*-values are presented in the tables below. Due to our small sample size, we present effect sizes calculated as standardized regression coefficients, or beta coefficients. We also present the mean and standard deviation (SD) of each measure of interest. **All comparisons presented in tables are relative to the Mask condition.**

3. Results

3.1. ASR performance

3.1.1. *Base ASR with no noise added to signal*

The WER for the Mask condition was lower than the WER for the control condition. Although this difference was not significant, there was a medium effect size. Of the speech strategies, only Clear speech had a significant effect on WER relative to the Mask condition. However, all three strategies resulted in a lower WER than the Mask condition; this effect was large for Clear speech and small for Loud and Slow speech.

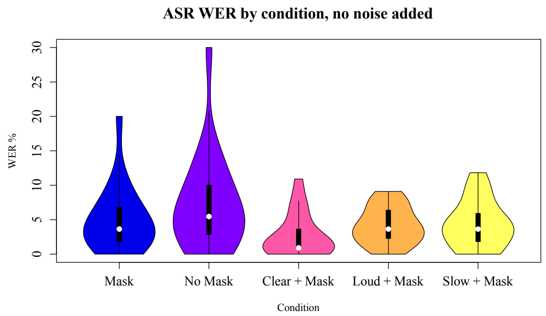

Figure 1: *Violin plots of the ASR WER for each condition.*

Table 1: *WER for base ASR analysis in each condition.*

	Beta	Effect size	p	WER (%) Mean	WER (%) SD
Mask	n/a	n/a	n/a	5.12	4.76
No Mask	.47	medium	.061	7.34	7.01
Clear+Mask	-.51	large	.044	2.73	3.11

	Beta	Effect size	p	WER (%)	
				Mean	SD
Loud+Mask	-.21	small	.409	4.16	2.66
Slow+Mask	-.16	small	.513	4.35	3.46

3.1.2. *ASR with noise added to signal at -1dB SNR*

When multi-talker babble was added to the signal to reduce a floor effect, there was no difference between the Mask and No Mask conditions. However, the Loud and Clear strategies significantly reduced WER relative to the Mask condition. Slow speech had a small effect on WER.

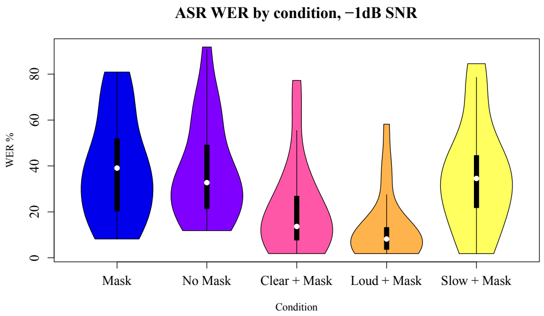

Figure 2: *Violin plots of the WER for each condition for samples mixed with noise at a -1dB SNR.*

Table 2: *WER for ASR analysis in each condition after noise mixing at -1dB SNR.*

	Beta	Effect size	p	WER (%)	
				Mean	SD
Mask	n/a	n/a	n/a	38.71	22.25
No Mask	-.05	none	.809	37.51	21.81
Clear+Mask	-.78	large	.001	20.81	20.18
Loud+Mask	-1.13	large	< .001	12.82	15.03
Slow+Mask	-.13	small	.477	35.22	23.41

3.2. Vowel space area (VSA)

VSA for the Clear condition was significantly larger than the Mask condition. VSA in the Mask condition was larger than VSA in the No Mask and Loud conditions, both with nonsignificant medium effect sizes. Slow speech had no effect on VSA relative to the Mask condition.

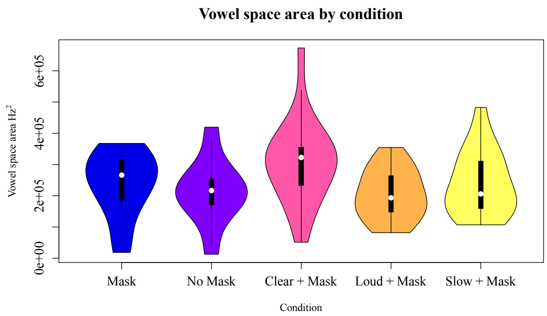

Figure 3: *Violin plots of the VSA for each condition.*

Table 3: *VSA for each condition.*

	Beta	Effect size	p	VSA (Hz^2)	
				Mean	SD
Mask	n/a	n/a	n/a	237196	102845
No Mask	-.218	small	.277	213385	101273
Clear+Mask	.651	large	.002	308400	133179
Loud+Mask	-.299	small	.137	204514	79427
Slow+Mask	.018	none	.927	239196	102649

3.3. Speaking rate

Speaking rate (measured as words per minute or WPM) during the Clear, Loud, and Slow conditions was significantly slower than during the Mask condition. Slow speech resulted in the slowest rate, followed by Clear speech. There was no difference in speaking rate between the No Mask and Mask conditions.

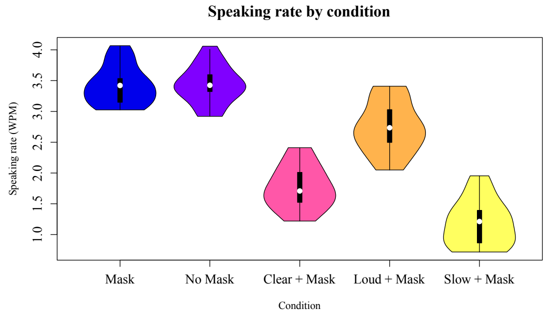

Figure 4: *Violin plots of the speaking rate for each condition.*

Table 4: *Speaking rate for each condition.*

	Beta	Effect size	p	Sp. rate (WPM)	
				Mean	SD
Mask	n/a	n/a	n/a	3.42	.31
No Mask	.01	none	.949	3.43	.30
Clear+Mask	-1.71	large	< .001	1.77	.34
Loud+Mask	-.71	large	< .001	2.74	.41
Slow+Mask	-2.32	large	< .001	1.18	.37

3.4. Speaking intensity

Speaking intensity in the Loud, Clear, and Slow conditions was significantly greater than the Mask condition. There was a slight increase in intensity and small effect size for the No Mask condition compared to the Mask Condition.

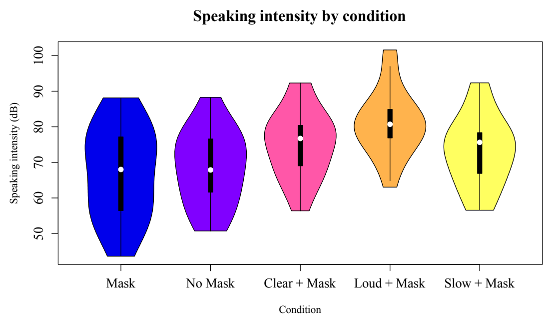

Figure 5: *Violin plots of the speech intensity for each condition.*

Table 5: *Speaking intensity for each condition.*

	Beta	Effect size	p	Intensity (dB) Mean	SD
Mask	n/a	n/a	n/a	67.11	12.84
No Mask	.12	small	.296	68.51	10.42
Clear+Mask	.71	large	< .001	75.25	9.53
Loud+Mask	1.23	large	< .001	81.16	9.34
Slow+Mask	.52	large	< .001	73.11	9.82

4. Discussion

4.1. No impact on ASR performance for healthy speakers wearing a mask

Our results suggest that there is no functional impact of wearing a KN95 mask on ASR performance for young, healthy speakers. It is important to note, however, that speakers during this study were not directly interacting with an ASR system; they may speak differently when receiving external feedback on their success with ASR. Moreover, ASR performance may differ in a less controlled environment or with spontaneous speech, and it could be worse under less ideal recording conditions.

4.2. Some evidence for natural compensation in speakers

Intelligibility enhancing adaptations to the mask were not evident in most of our acoustic measures of speech, which is not surprising given how well the ASR performed during the Mask condition. Furthermore, because our participants were not directly interacting with another person or ASR system, the lack of communicative intent in the experimental paradigm could have impacted our findings [32]. However, even in the absence of external feedback from an ASR system, we would expect speakers to adjust their speech based on internal feedback mechanisms, similar to the effects seen in adaptation paradigms [33], [34]. For example, when speaking in noisy environments, speakers tend to increase their vocal intensity [34]. Nonetheless, we reported no significant differences in speaking rate, VSA, nor intensity when comparing the Mask and No Mask conditions. However, we did find evidence that speakers were hyperarticulating their speech (as indexed by a medium effect size for increased VSA), which warrants future work with a larger sample size. Moreover, we found a slight decrease in speaking intensity for the Mask condition relative to the No Mask condition. This reduction was nonsignificant and runs contrary to expectations of a large decrease in intensity for the Mask condition. However, it is possible that our participants increased vocal effort to effectively compensate for the mask's attenuating effect on speech.

4.3. Explicit compensatory strategies improve ASR performance for speakers wearing masks

Interestingly, while speakers did not automatically engage in compensatory strategies in the Mask condition, explicitly instructing speakers to employ Clear or Loud speech significantly reduced ASR WER. Clear speech had the greatest impact on ASR performance, as it produced substantial change even in the base ASR condition. Slow speech did not significantly improve ASR performance, although it did result in a slight reduction of WER (as indexed by a small effect size). Moreover, the gains in ASR performance were accompanied by acoustic changes that indicate that speakers were indeed producing speech that was clear (increased VSA), loud (increased intensity), and slow (decreased rate). This advantage of compensatory strategies may have important implications for speakers with reduced intelligibility, such as people with speech impairment or non-native English speakers. ASR systems tend to perform more poorly for speakers with reduced intelligibility [16], and the acoustic effect of masks may degrade the already impoverished speech signal of these speakers. Moreover, speakers with impaired speech or language systems may be less able to automatically adapt their speaking to overcome the effects of wearing a mask due to their underlying impairment or reduced perceptual awareness of their speech changes [33]. Thus, in situations where the ASR user is wearing a mask and presents with lower speech intelligibility, using Clear or Loud speech may significantly improve the accuracy of the ASR system. Although not tested in our study, these strategies could presumably improve accessibility and communicative participation for speakers who use ASR for such purposes (e.g., using Google Assistant to adjust the lights or to repeat utterances more clearly to a communication partner).

4.4. Future Directions

Future work will investigate the impact of compensatory speech strategies on perceptual speech intelligibility of mask-wearers; such work may provide guidance for speakers wearing masks and their communication partners, including non-native English speakers and listeners with hearing loss. Future work is also needed to assess the impact of mask-wearing on ASR performance in a more ecologically valid setting. In addition, further research should directly assess the impacts of mask-wearing on ASR performance in individuals with speech impairments. The acoustic effects of KN95 masks also merit deeper investigation, as they have largely been excluded from studies. Lastly, this study investigated vowel space area, speech intensity, and speaking rate as indicators of compensation; however, speakers may employ other measures of natural compensation, such as increased consonant contrasts or prosodic changes, including altered pitch or phrase boundary markers.

5. Conclusion

Our study found that wearing a KN95 mask does not affect commercial ASR performance at the sentence level for young, healthy speakers. Additionally, while we found little evidence of natural compensation for the mask, there were significant improvements in ASR performance when speakers employed compensatory speech strategies. This work highlights the benefits of using speech strategies on ASR performance when the end-user is wearing a mask. Taken together, our results have implications for healthy speakers and potentially greater implications for speakers with impairments that inhibit their success with ASR.

6. Acknowledgements

This work was supported by NIH-NIDCD under Grants K24DC016312, T32DC000038, and F31DC019016.

7. References

[1] OSHA 1910.134, "Personal Protective Equipment: Respiratory Protection." 2011, [Online]. Available: http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=9780.

[2] CDC, "Coronavirus disease 2019 (COVID-19). Recommendation for cloth face covers.," Atlanta, GA, 2020.
[3] K. A. Fisher *et al.*, "Factors Associated with Cloth Face Covering Use Among Adults During the COVID-19 Pandemic — United States, April and May 2020," *MMWR. Morb. Mortal. Wkly. Rep.*, vol. 69, no. 28, pp. 933–937, Jul. 2020, doi: 10.15585/mmwr.mm6928e3.
[4] M. Abkarian, S. Mendez, N. Xue, F. Yang, and H. A. Stone, "Speech can produce jet-like transport relevant to asymptomatic spreading of virus," *Proc. Natl. Acad. Sci. U. S. A.*, 2020, doi: 10.1073/pnas.2012156117.
[5] L. Morawska *et al.*, "Size distribution and sites of origin of droplets expelled from the human respiratory tract during expiratory activities," *J. Aerosol Sci.*, 2009, doi: 10.1016/j.jaerosci.2008.11.002.
[6] S. Asadi, A. S. Wexler, C. D. Cappa, S. Barreda, N. M. Bouvier, and W. D. Ristenpart, "Aerosol emission and superemission during human speech increase with voice loudness," *Sci. Rep.*, 2019, doi: 10.1038/s41598-019-38808-z.
[7] R. Saeidi, I. Huhtakallio, and P. Alku, "Analysis of face mask effect on speaker recognition," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 08-12-Sept, pp. 1800–1804, 2016, doi: 10.21437/Interspeech.2016-518.
[8] R. Saeidi, T. Niemi, H. Karppelin, J. Pohjalainen, T. Kinnunen, and P. Alku, "Speaker recognition for speech under face cover," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2015-Janua, pp. 1012–1016, 2015.
[9] M. Vojnović and M. Mijić, "The influence of the oxygen mask on long-time spectra of continuous speech," *J. Acoust. Soc. Am.*, vol. 102, no. 4, pp. 2456–2458, 1997, doi: 10.1121/1.421021.
[10] A. Goldin, B. Weinstein, and N. Shiman, "How do medical masks degrade speech reception?," *Hear. Rev.*, no. May, pp. 8–9, 2020.
[11] S. R. Atcherson *et al.*, "The effect of conventional and transparent surgical masks on speech understanding in individuals with and without hearing loss," *J. Am. Acad. Audiol.*, vol. 28, no. 1, pp. 58–67, 2017, doi: 10.3766/jaaa.15151.
[12] S. Asadi, C. D. Cappa, S. Barreda, A. S. Wexler, N. M. Bouvier, and W. D. Ristenpart, "Efficacy of masks and face coverings in controlling outward aerosol particle emission from expiratory activities," *Sci. Rep.*, 2020, doi: 10.1038/s41598-020-72798-7.
[13] P. Bottalico, S. Murgia, G. E. Puglisi, A. Astolfi, and K. I. Kirk, "Effect of masks on speech intelligibility in auralized classrooms," *J. Acoust. Soc. Am.*, vol. 148, no. 5, pp. 2878–2884, Nov. 2020, doi: 10.1121/10.0002450.
[14] N.-C. C. Ristea and R. T. Ionescu, "Are you wearing a mask? Improving mask detection from speech using augmentation by cycle-consistent GANs," *Proc. INTERSPEECH 2020 – 21st Annu. Conf. Int. Speech Commun. Assoc.*, pp. 2102–2106, Jun. 2020, doi: 10.21437/interspeech.2020-1329.
[15] M. Markitantov, D. Dresvyanskiy, D. Mamontov, H. Kaya, W. Minker, and A. Karpov, "Ensembling End-to-End Deep Models for Computational Paralinguistics Tasks: ComParE 2020 Mask and Breathing Sub-Challenges," *Proc. INTERSPEECH 2020 – 21st Annu. Conf. Int. Speech Commun. Assoc.*, pp. 2072–2076, 2020, doi: 10.21437/Interspeech.2020-2666.
[16] L. J. Ferrier, H. C. Shane, H. F. Ballard, T. Carpenter, and A. Benoit, "Dysarthric Speakers Intelligibility and Speech Characteristics in Relation to Computer Speech Recognition," *Augment. Altern. Commun.*, vol. 11, no. 3, pp. 165–175, 1995, doi: 10.1080/07434619512331277289.
[17] A. J. Palmiero, D. Symons, J. W. Morgan, and R. E. Shaffer, "Speech intelligibility assessment of protective facemasks and air-purifying respirators," *J. Occup. Environ. Hyg.*, vol. 13, no. 12, pp. 960–968, 2016, doi: 10.1080/15459624.2016.1200723.
[18] L. J. Radonovich, R. Yanke, J. Cheng, and B. Bender, "Diminished speech intelligibility associated with certain types of respirators worn by healthcare workers," *J. Occup. Environ. Hyg.*, 2010, doi: 10.1080/15459620903404803.
[19] C. Llamas, P. Harrison, D. Donnelly, and D. Watt, "Effects of Different Types of Face Coverings on Speech Acoustics and Intelligibility," *York Pap. Linguist. Ser. 2*, no. 9, pp. 80–104, 2008.
[20] J. Lam and K. Tjaden, "Intelligibility of Clear Speech: Effect of Instruction," *J. Speech, Lang. Hear. Res.*, vol. 56, no. 5, pp. 1429–1440, 2013, doi: 10.1044/1092-4388(2013/12-0335).
[21] J. C. Krause and L. D. Braida, "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," *J. Acoust. Soc. Am.*, vol. 112, no. 5, pp. 2165–2172, Nov. 2002, doi: 10.1121/1.1509432.
[22] C. Fox, L. Ramig, M. Ciucci, S. Sapir, D. McFarland, and B. Farley, "The Science and Practice of LSVT/LOUD: Neural Plasticity-Principled Approach to Treating Individuals with Parkinson Disease and Other Neurological Disorders," *Semin. Speech Lang.*, vol. 27, no. 4, pp. 283–299, Nov. 2006, doi: 10.1055/s-2006-955118.
[23] K. Tjaden, J. E. Sussman, and G. E. Wilding, "Impact of Clear, Loud, and Slow Speech on Scaled Intelligibility and Speech Severity in Parkinson's Disease and Multiple Sclerosis," *J. Speech, Lang. Hear. Res.*, vol. 57, no. 3, pp. 779–792, Jun. 2014, doi: 10.1044/2014_JSLHR-S-12-0372.
[24] K. M. Yorkston, M. Hakel, D. R. Beukelman, and S. Fager, "Evidence for effectiveness of treatment of loudness, rate, or prosody in dysarthria: A systematic review," *J. Med. Speech. Lang. Pathol.*, vol. 15, no. 2, 2007.
[25] K. Tjaden, J. Lam, and G. Wilding, "Vowel acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions," *J. Speech, Lang. Hear. Res.*, vol. 56, no. 5, pp. 1485–1502, 2013, doi: 10.1044/1092-4388(2013/12-0259).
[26] J. Lam, K. Tjaden, and G. Wilding, "Acoustics of Clear Speech: Effect of Instruction," *J. Speech Lang. Hear. Res.*, vol. 55, no. 6, p. 1807, 2012, doi: 10.1044/1092-4388(2012/11-0154).
[27] K. Yorkston, D. Beukelman, and M. Hakel, "Speech Intelligibility Test (SIT) for Windows." 2007.
[28] J. R. Green, I. S. B. Nip, E. M. Wilson, A. S. Mefferd, and Y. Yunusova, "Lip Movement Exaggerations During Infant-Directed Speech," *J. Speech, Lang. Hear. Res.*, vol. 53, no. 6, pp. 1529–1542, Dec. 2010, doi: 10.1044/1092-4388(2010/09-0005).
[29] J. R. Green, D. R. Beukelman, and L. J. Ball, "Algorithmic Estimation of Pauses in Extended Speech Samples of Dysarthric and Typical Speech.," *J. Med. Speech. Lang. Pathol.*, vol. 12, no. 4, pp. 149–154, Dec. 2004, Accessed: Mar. 28, 2019. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20628555.
[30] L. Fontan, J. Tardieu, P. Gaillard, V. Woisard, and R. Ruiz, "Relationship between speech intelligibility and speech comprehension in babble noise," *J. Speech, Lang. Hear. Res.*, vol. 58, no. 3, pp. 977–986, 2015, doi: 10.1044/2015_JSLHR-H-13-0335.
[31] D. R. Mccloy, "Vowel normalization and plotting with the phonR package," *Tech. Reports UW Linguist. Phonetics Lab.*, 2012.
[32] V. Hazan and R. Baker, "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," *J. Acoust. Soc. Am.*, vol. 130, no. 4, pp. 2139–2152, Oct. 2011, doi: 10.1121/1.3623753.
[33] V. M. Villacorta, J. S. Perkell, and F. H. Guenther, "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," *Cit. J. Acoust. Soc. Am.*, vol. 122, p. 2306, 2007, doi: 10.1121/1.2773966.
[34] J. C. Junqua, "The lombard reflex and its role on human listeners and automatic speech recognizers," *J. Acoust. Soc. Am.*, vol. 93, no. 1, pp. 510–524, Jan. 1993, doi: 10.1121/1.405631.