Exploring listeners' speech rate preferences

Olympia Simantiraki[1], *Martin Cooke*[2]

[1]Language and Speech Laboratory, University of the Basque Country, Vitoria, Spain
[2]Ikerbasque (Basque Science Foundation)

olympia.simantiraki@ehu.eus, m.cooke@ikerbasque.org

Abstract

Fast speech may reduce intelligibility, but there is little agreement as to whether listeners benefit from slower speech in noisy conditions. The current study explored the relationship between speech rate and masker properties using a listening preference technique in which participants were able to control speech rate in real time. Spanish listeners adjusted speech rate while listening to word sequences in quiet, in stationary noise at signal-to-noise ratios of 0, +6 and +12 dB, and in modulated noise for 5 envelope modulation rates. Following selection of a preferred rate, participants went on to identify words presented at that rate. Listeners favoured faster speech in quiet, chose increasingly slower rates in increasing levels of stationary noise, and showed a preference for speech rates that led to a contrast with masker envelope modulation rates. Participants showed distinct preferences even when intelligibility was near ceiling levels. These outcomes suggest that individuals attempt to compensate for the decrement in cognitive resources availability in more adverse conditions by reducing speech rate and are able to exploit differences in modulation properties of the target speech and masker. The listening preference approach provides insights into factors such as listening effort that are not measured in intelligibility-based metrics.

Index Terms: speech perception, speech rate, stationary noise, modulated noise, listener preferences

1. Introduction

Synthetic and recorded speech form an increasing part of our everyday listening experience, and much of our exposure to these forms of speech occurs in potentially noisy settings such as on public transport, in the classroom or workplace, while driving, and in our homes. Optimising speech output to ensure that salient information is correctly and effortlessly received is an important problem for the designers of applications that make use of the speech modality. The current study uses a listener preference paradigm to explore the effects of changes in speech rate in noisy conditions.

Most of the focus in adapting speech output to challenging listening conditions has been on intelligibility, and specifically on enhancing intelligibility by modifying speech prior to presentation e.g. [1, 2, 3]. Modification techniques applied to natural [4, 5] and synthetic [6] speech have been shown to be capable of substantial intelligibility improvements for speech presented in noisy conditions, with gains equivalent to increasing the volume by more than 5 dB [7].

However, it is questionable as to whether intelligibility-enhancement should be the sole motivation for speech modification, for a number of reasons. First, the typically negative signal-to-noise ratios (SNRs) at which substantial intelligibility gains are observed are not representative of those experienced in everyday listening conditions, where average SNRs are less adverse, at around 5 dB [8]. Consequently, it is likely that the gains observed in realistic conditions will be more modest than laboratory studies suggest. Second, a focus on intelligibility alone ignores the additional cognitive demands – listening effort – that speech output makes on listeners over and above the desire to comprehend meaning [9, 10, 11]. The cognitive implications of intelligibility-enhancing speech modifications have not been assessed extensively, and it is conceivable that modifications that improve intelligibility lead to speech that is more demanding to process. Similarly, studies into how talkers adapt their speech as a function of context (e.g. clear speech [12]; speech produced in noise [13]; infant-directed speech [14]) have demonstrated consistent patterns of change for parameters such as intensity, vowel space, spectral tilt, and speech rate. While the impact of these changes on intelligibility has been studied (e.g., see review in [15]), investigations of their effect on listening effort are rare. However, emerging findings indicate that modifications can indeed affect effort. For example, by using pupillometry as a physiological proxy for listening effort it has been shown that poor quality synthetic speech is more effortful to process [16] and that existing intelligibility enhancement techniques can also reduce cognitive load [17].

The current study focuses on one feature of speech delivery that can be expected to impact both intelligibility and listening effort: speech rate. Fast speech has been found to disrupt intelligibility of both natural speech [18, 19] and synthetic speech [20, 21]. However, there is little agreement on whether a slow speech rate is beneficial [22, 23, 24, 25, 26]. One recent study [26] found no intelligibility gains for linearly-elongated speech when presented in stationary noise, but significant gains for the same speech in the presence of both competing speech or speech-shaped noise whose envelope was modulated by that of competing speech. However, it was unclear whether the benefit was due to the net availability of more phonetic information due to the dips in the masker, or to a difference in modulation rates between the target and masker speech. One of the goals of the current work is to examine the effect of speech rate changes in different noise conditions, including those involving variations in the modulation rate of the masker.

In this study we present an experimental approach whose goal is to simultaneously measure both intelligibility and any supra-intelligibility impact of speech modifications. The paradigm inverts the traditional technique of asking listeners to passively judge the intelligibility of preselected speech-in-noise materials. Instead, listeners are able to control the parameter under study – here, speech rate – during an adaptation phase in each test trial, prior to an identification phase in which speech is presented at the preferred rate. In this way listener preferences and their impact are measured directly, rather than through opinion scores, using an easy-to-understand procedure akin to adjusting the volume of a device using a remote control. The closest study to the current work [27] used an on-line speech dilation technique [28] to allow listeners to modify speech rate in real-time using an on-screen slider bar. Our study extends this work by investigating the impact of different masker types and masker modulation rates on listener preferences, while also allowing listeners to change speech rate in the faster direction as well as slowing it down.

The main questions addressed in the current study are: do listeners show speech rate preferences that lead to intelligible speech; are these rates consistent across the listener cohort; do speech rate preferences show patterns that are independent of intelligibility; and how do speech rate preferences change in challenging conditions?

2. Methods

Listeners changed speech rate in an open-ended adjustment phase, followed by a fixed length test phase (sec. 2.3). In separate conditions, listeners adjusted and identified speech in quiet, in speech-shaped noise, and in speech-modulated noise (sec. 2.2).

2.1. Participants

Eighteen native Spanish listeners (15 females) aged 18-23 (mean 19.9, $SD = 1.4$) participated in the experiment. All passed an audiological screening with a hearing level better than 25 dB at frequencies in the range $125 - 8000$ Hz in both ears. Listeners were paid for their participation.

2.2. Stimuli

Speech material consisted of sequences of isolated Spanish words spoken by a female talker, drawn from an open source 3968-word corpus of the most frequent Spanish words of up to three syllables [29]. Twenty-two different speech rates were available for listeners to choose covering the range from 2.5 times slower than the original to 2.5 times faster, with speech rates located at equally-spaced points on a multiplicative inverse scale ($\frac{1}{[0.4:0.1:2.5]}$). Since the target speech material was read speech and hence not as fast as casual speech, we determined on the basis of informal listening tests to deploy 15 steps with the speech rate faster than the original, one at with original rate, and 6 with rates slower than the original. Linear elongation/compression was employed, and all rate modifications were carried out using the STRAIGHT toolbox [30]. Words were independently normalised to have equal root mean square level after rate modifications, with 20 ms half-Hamming ramps applied to reduce onset/offset transients.

Stimuli were presented in quiet and in 8 additive noise masking conditions: speech-shape noise (SSN) at SNRs of 0, +6 and +12 dB, and speech modulated noise (SMN) for 5 envelope modulation rates, mixed with speech at +6 dB SNR. Maskers were based on concatenated Spanish sentences from the Sharvard corpus [31], spoken by a female talker. Maskers were unrelated to the target speech stimuli described above. The SSN masker resulted from passing random uniform noise through a filter with the long-term spectrum of the concatenated sentences. The SMN masker was generated by multiplying the SSN masker by the instantaneous envelope of the concatenated sentences. In addition to the original rate, envelope modulation rates 1.4 and 2.5 times faster, and 1.7 and 2.5 times slower than the original were tested (rates of 1.4 and 1.7 correspond to equidistant steps from the original rate on the 22-point scale used here). In the masked conditions SNRs were computed by concatenating stimulus words without gaps.

2.3. Procedure

The experiment was blocked into the 9 conditions described above. Across participants, block order was counterbalanced using a Latin square design. Each block contained 22 trials. A trial started with a speech rate randomly-chosen from the 22 available, and consisted of an adjustment phase followed by a test phase. During the adjustment phase, words were presented in a randomised order with 500 ms of intervening silence. Participants were instructed to choose an optimal value of speech rate that allowed them to recognise as many words as possible. Participants were able to control speech rate using the up/down arrow keys to speed up or slow down. Changes in speech rate occurred on the next word presented. The adjustment phase continued for as long as participants required (which averaged 7.04 s; $SD = 5.59$). Having chosen a preferred speech rate, participants were able to proceed to the test phase by clicking an on-screen button. Participants were not allowed to proceed to the test phase until at least five seconds of the adjustment phase had elapsed. In the test phase, participants were presented with words spoken at the speech rate chosen in the adjustment phase, under the same experimental condition as they had experienced during the adjustment phase. Participants typed their responses into an on-screen text input box. During each test phase of a single trial, five test words were presented consecutively and across conditions no word was repeated. In total, listeners responded to 990 unique words (9 conditions x 22 trials x 5 test items) during the experiment.

Prior to the main experiment, listeners were given written guidance which encouraged them to think of the task in the same way as choosing an appropriate volume for the television: too quiet makes it difficult to understand the words, and too loud is uncomfortable. Then they underwent a familiarisation phase consisting of 5 trials in each of Quiet, SSN and SMN (at a single SNR for the masked conditions). The entire experiment lasted around two hours, and participants were able to take a break between each block. Stimuli were presented through Sennheiser HD380 headphones at a fixed presentation level while listeners were seated in sound-attenuating booths in a purpose-build speech perception laboratory at the University of the Basque Country (Alava Campus).

3. Results

3.1. Speech rate preferences

Figure 1 plots speech rate preferences, intelligibility and the time spent in the adjustment phase for the 9 conditions of the study. In quiet, listeners preferred to listen to speech 1.2 times faster than the original rate, and at 97.7%, word scores were close to ceiling. Listeners spent 5.4 s during the adjustment phase in this condition, close to the 5 s minimum permitted. Compared to quiet, in noise listeners selected slower speech rates and spent longer on adjustment. Even at the various 'optimal' speech rates (i.e. those most-frequently selected) for the different masking conditions, listeners did not achieve intelligibility scores as high as those in the Quiet condition. Within each masker type (SMN, SSN) adjustment time was longer for conditions resulting in lower intelligibility. However this was not true across masker types; for example, less time was spent adjusting in the 0 dB SNR condition for the SSN masker than in some of the SMN conditions even though intelligibility was substantially lower in the 0 dB SNR SSN condition.

Separate one-way within-subjects ANOVAs conducted to compare the effect of condition on each of the three measure-

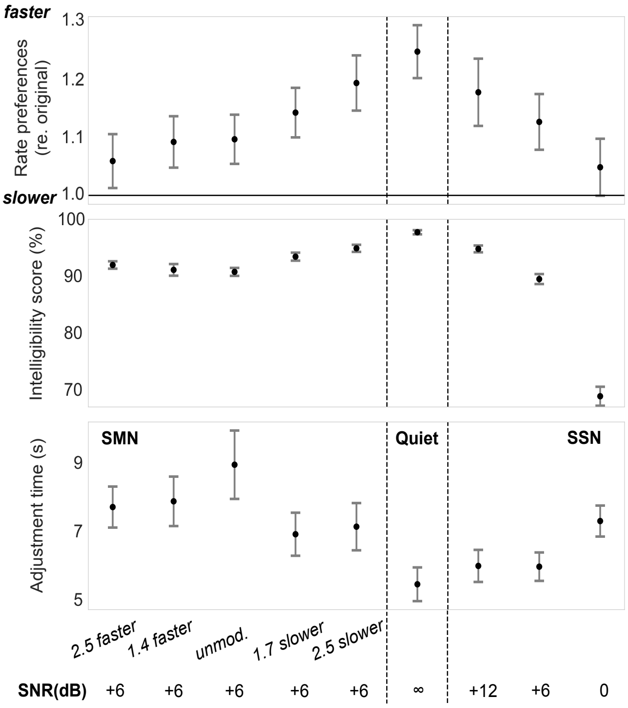

Figure 1: *Speech rate preferences (upper plot), intelligibility scores (middle plot) and adjustment time (lower plot) are depicted. Dashed vertical lines separate the SMN, Quiet, and SSN conditions. The solid horizontal line in the upper plot indicates the original speech rate. Error bars represent ± 1 SE.*

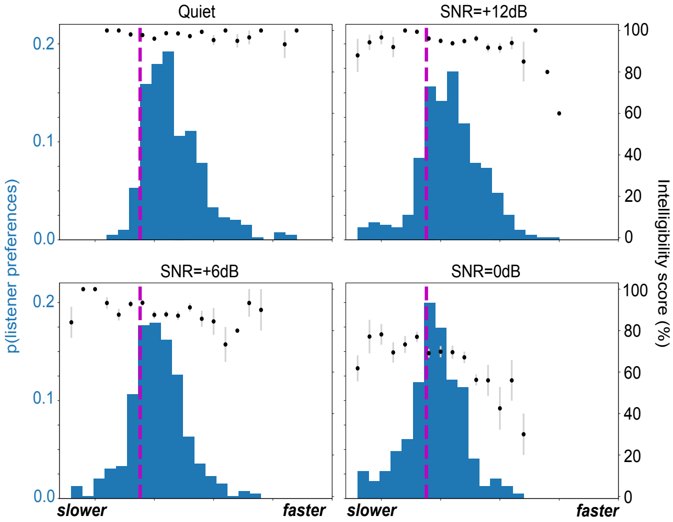

Figure 2: *Probability of each speech rate value preference (histogram, left-axis) for the Quiet and SSN conditions, along with the percentage of words recalled correctly (black dots, right-axis). Error bars represent ± 1 SE. The dashed line denotes the step that corresponds to the speech rate of the original speech signal.*

ments indicated significant main effects on speech rate preferences [$F(8, 136) = 9.1, p < .001, \eta^2 = .09$], adjustment time [$F(8, 136) = 6.1, p < .001, \eta^2 = .14$] and intelligibility [$F(8, 136) = 124.7, p < .001, \eta^2 = .83$]. Post-hoc comparisons using the Tukey HSD test indicated that all conditions, relative to the Quiet baseline, resulted in significantly lower intelligibility, longer adjustment time (except for SSN at +6 and +12 dB SNR), and slower preferred speech rate (except SMN at the 2.5 times slower modulation rate). For the SSN conditions, these comparisons showed increasingly higher intelligibility with increasing SNR and for the adverse noise level significantly higher adjustment time and slower rate compared to the less noisy conditions. For the SMN conditions, the two slower masker modulation rates led to significantly higher intelligibility than the faster modulation rates. Significantly less time was needed to adjust speech rate when the masker's modulation rate was different than the original. Adjustment time in the face of modulated maskers was longer than for the stationary maskers apart from the least adverse SSN condition. Finally, listeners preferred significantly faster speech when the masker modulation rate was slow, with a tendency towards the converse when the masker modulation rate was fast.

3.2. Listener preferences and intelligibility

The probability with which each of the 22 permitted speech rate values was preferred by listeners, along with the percentage of keywords correctly recalled at that speech rate, is presented in Fig. 2 for the Quiet and SSN conditions and in Fig. 3 for the Quiet and SMN masker conditions (the Quiet condition is repeated for convenience). At the most adverse SSN condition

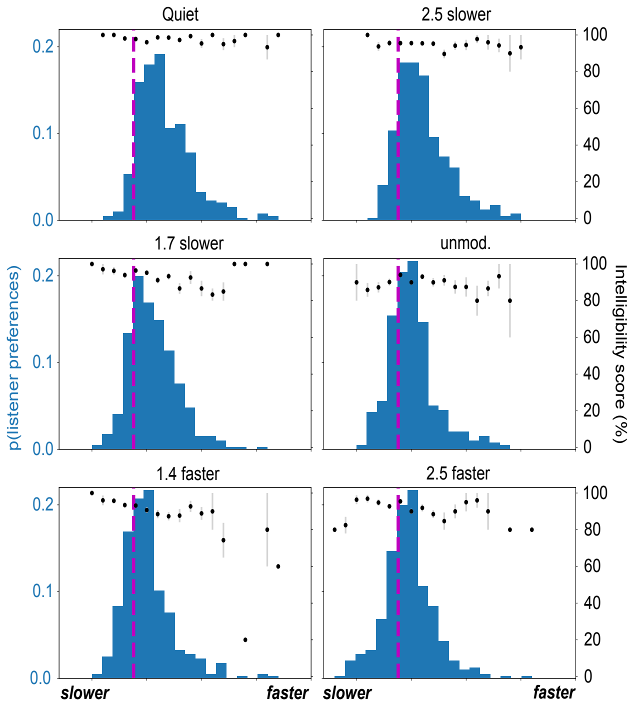

Figure 3: *As Fig. 2 but for the Quiet and SMN conditions.*

(0 dB SNR) it is notable that those listeners who chose faster rates produced lower intelligibility scores, but in general intelligibility scores were relatively uniform across speech rates, though not necessarily at ceiling levels. Nevertheless, listeners showed distinct preferences for certain speech rates as manifested by the relatively sharply-peaked preference distributions.

Inter-rater reliability was determined using two-way intraclass correlation [32] to assess the degree that listeners provided consistency in their mean speech rate choices across conditions, using the *icc* function of *irr* package in R. The resulting intraclass correlation value of .879 was in the 'excellent' range [33], indicating that listeners had a high degree of agreement in selecting preferred speech rates.

3.3. Initial speech rate preference versus adjustment time

Repeated-measures correlation via the *rmcorr* package in R [34] showed that the initial speech rate value of each trial was positively correlated with the final preference (i.e. a fast initial rate tended to lead to fast speech at the end of adjustment, and vice versa) [$p < .001$] and negatively correlated with adjustment time [$p < .001$]. In other words, when the initial speech rate was far from the 'optimal' value (considered as the mean value chosen by the cohort of listeners at the end of adjustment), listeners tended not to tune it all the way to this value.

4. Discussion

The current study found distinct speech rate preferences that showed up even with intelligibility at ceiling. Such preferences reveal supra-intelligibility aspects of speech rate, suggesting that in stationary noise listeners prefer a slower speech rate as noise level increases, while for fluctuating noise they prefer faster speech when masker modulation rate is slow and vice versa. The preferred speech rate in quiet was faster than any of the masking conditions. These findings are in line with the real-time speech rate modification study of [27], whose listeners chose to expand speech at adverse SNRs in the face of a 4-talker babble masker even though such preferences did not improve intelligibility.

For stationary noise, it has been argued [26] that linear elongation of speech resulting from a slower speech rate is not beneficial to intelligibility because it merely leads to elongation of those 'glimpses' of speech that escape masking rather than revealing additional speech information. Our findings support this claim: in general, intelligibility did not improve for listeners who chose slower speech (Fig. 2). The fact that listeners preferred slower rates in more adverse stationary noise might indicate a desire to reduce listening effort, or could reflect an attempt to reproduce typical speech rates experienced by participants in real-world noisy conditions which are characterised by slower speech [35].

Concerning the speech-modulated noise conditions, listeners tended to prefer a target speech rate that contrasted with that of the masker. The presence of similar fluctuation rates would be expected to lead to difficulties in pulling out target speech from a masker [36]. Differences in modulation rate might act as a cue for segregating the two signals, a possibility supported by a finding that intelligibility improves when the fluctuation rates of target speech and a babble noise masker are mismatched [37]. A contrast in modulation rates might help in a number of ways. One is to allow a target to be tracked through time by sequential grouping of those speech fragments with similar rates. Indeed, some listeners in [27] reported that their speech rate choices helped them to track the target speaker. A complementary possibility is that listeners manipulate speech rate in order to promote energetic masking release. For example, a faster rate potentially allows more evidence of the target to 'fit' in the longer temporal dips of a masker with slow modulations. There is some evidence that talkers adopt just such a strategy when 'listening-while-speaking' [38].

Listeners spent more time adjusting the target speech rate in the presence of temporally-modulated noise than in the other conditions. The modulated nature of the masker allows varying amounts of target speech energy to be audible at different points, and it is possible that this causes additional cognitive load for the listener and make it harder to predict when to listen. In [39] more perceived effort was reported for noises with a high degree of temporal variation at a relatively high SNR ($+10$ dB). In [40], at low SNRs listeners rated stationary masking as more effortful than fluctuating noise, but the difference between the two types of masker was reduced or eliminated with increasing SNR, leading to the suggestion that peaks in the fluctuating masker might have a negative impact on listeners in less noisy conditions.

Finally, we note that although listeners preferred reduced speech rates in adverse conditions, in all conditions the mean rate chosen was faster than the original speech (Fig. 1). This is most likely due to the use of read speech, which is typically somewhat slower than normal or casual speech [41].

5. Conclusions

The current study explored listeners' speech rate preferences in stationary and temporally-modulated noise. Listeners chose slower speech in noise than in quiet, with greater reductions as the level of the stationary masker increased. For modulated maskers listeners tended to prefer speech rates that led to a contrast with the modulation rate of the masker. Speech rate preferences did not simply reflect choices that led to increased intelligibility. These findings suggest that listener preferences provide information about the processing of speech over and above that measured by intelligibility, and that such preferences may result from a desire to reduce the cognitive effort of understanding speech in adverse conditions.

6. Acknowledgements

Olympia Simantiraki was funded by the European Commission under the Marie Curie European Training Network ENRICH (675324).

7. References

[1] M. D. Skowronski and J. G. Harris, "Applied principles of clear and Lombard speech for automated intelligibility enhancement in noisy environments," *Speech Communication*, vol. 48, no. 5, pp. 549–558, 2006.

[2] H. Brouckxon, W. Verhelst, and B. D. Schuymer, "Time and Frequency Dependent Amplification for Speech Intelligibility Enhancement in Noisy Environments," in *Proc. Interspeech*, vol. 9, 2008, pp. 557–560.

[3] Y. Tang, B. Fazenda, and T. Cox, "Automatic Speech-to-Background Ratio Selection to Maintain Speech Intelligibility in Broadcasts Using an Objective Intelligibility Metric," *Applied Sciences*, vol. 8, p. 59, 2018.

[4] T. Zorila, V. Kandia, and Y. Stylianou, "Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression," in *Proc. Interspeech*, 2012, pp. 635–638.

[5] H. Schepker, J. Rennies, and S. Doclo, "Improving speech intelligibility in noise by SII-dependent preprocessing using frequency-dependent amplification and dynamic range compression," in *Proc. Interspeech*, 2013, pp. 3577–3581.

[6] C. Valentini-Botinhao, J. Yamagishi, and S. King, "Mel cepstral coefficient modification based on the Glimpse Proportion measure for improving the intelligibility of HMM-generated synthetic speech in noise," in *Proc. Interspeech*, 2012, pp. 631–634.

[7] M. Cooke, C. Mayo, and C. Valentini-Botinhao, "Intelligibility-enhancing speech modifications: the Hurricane Challenge," in *Proc. Interspeech*, 2013, pp. 3552–3556.

[8] K. Smeds, F. Wolters, and M. Rung, "Estimation of signal-to-noise ratios in realistic sound scenarios," *J. Am. Acad. Audiol.*, vol. 26, pp. 183–196, 2015.

[9] J. Rönnberg, M. Rudner, T. Lunner, and A. A. Zekveld, "When cognition kicks in: Working memory and speech understanding in noise," *Noise and Health*, vol. 12, pp. 263–269, 2010.

[10] C. Pals, A. Sarampalis, and D. Baskent, "Listening effort with cochlear implant simulations," *J. Speech Hearing Language Res.*, vol. 56, pp. 1075–1084, 2013.

[11] B. Hornsby, "The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands," *Ear Hearing*, vol. 34, pp. 523–534, 2013.

[12] R. M. Uchanski, "Clear speech," in *The Handbook of Speech Perception*, 2005, pp. 207–235.

[13] W. V. Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow, and M. A. Stokes, "Effects of noise on speech production: Acoustic and perceptual analyses," *J. Acoust. Soc. Am.*, vol. 84, no. 3, pp. 917–928, 1988.

[14] J. Y. Song, K. Demuth, and J. Morgan, "Effects of the acoustic properties of infant-directed speech on infant word recognition," *J. Acoust. Soc. Am.*, vol. 128, no. 1, pp. 389–400, 2010.

[15] M. Cooke, S. King, M. Garnier, and V. Aubanel, "The listening talker: A review of human and algorithmic context-induced modifications of speech," *Computer Speech and Language*, vol. 28, pp. 543–571, 2014.

[16] A. Govender and S. King, "Using Pupillometry to Measure the Cognitive Load of Synthetic Speech," in *Proc. Interspeech*, 2018, pp. 2838–2842.

[17] O. Simantiraki, M. Cooke, and S. King, "Impact of Different Speech Types on Listening Effort," in *Proc. Interspeech*, 2018, pp. 2267–2271.

[18] G. Fairbanks and F. Kodman, "Word intelligibility as a function of time compression," *J. Acoust. Soc. Am.*, vol. 29, pp. 636–644, 1957.

[19] N. J. Versfeld and W. A. Dreschler, "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners," *J. Acoust. Soc. Am.*, vol. 111, no. 1, pp. 401–408, 2002.

[20] J. Lebeter and S. Saunders, "The effects of time compression on the comprehension of natural and synthetic speech," *Working Papers of the Linguistics Circle*, vol. 20, no. 1, pp. 63–81, 2010.

[21] C. Valentini-Botinhao, M. Toman, M. Pucher, D. Schabus, and J. Yamagishi, "Intelligibility Analysis of Fast Synthesized Speech," in *Proc. Interspeech*, 2014, pp. 2922–2926.

[22] E. M. Adams and R. E. Moore, "Effects of speech rate, background noise, and simulated hearing loss on speech rate judgment and speech intelligibility in young listeners," *J. Am. Acad. Audiol.*, vol. 20, pp. 28–39, 2009.

[23] E. M. Adams, S. Gordon-Hickey, H. Morlas, and R. Moore, "Effect of Rate-Alteration on Speech Perception in Noise in Older Adults With Normal Hearing and Hearing Impairment," *Am. J. Audiol.*, vol. 21, no. 1, pp. 22–32, 2012.

[24] Y. Nejime and B. C. J. Moore, "Evaluation of the effect of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss," *J. Acoust. Soc. Am.*, vol. 103, no. 1, pp. 572–576, 1998.

[25] M. Cooke, C. Mayo, and J. Villegas, "The contribution of durational and spectral changes to the Lombard speech intelligibility benefit," *J. Acoust. Soc. Am.*, vol. 135, no. 2, pp. 874–883, 2014.

[26] M. Cooke and V. Aubanel, "Effects of linear and nonlinear speech rate changes on speech intelligibility in stationary and fluctuating maskers," *J. Acoust. Soc. Am.*, vol. 141, no. 6, pp. 4126–4135, 2017.

[27] J. S. Novak III and R. V. Kenyon, "Effects of User Controlled Speech Rate on Intelligibility in Noisy Environments," in *Proc. Interspeech*, 2018, pp. 1853–1857.

[28] J. S. Novak III, A. Tandon, J. Leigh, and R. V. Kenyon, "Networked On-Line Audio Dilation," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. Association for Computing Machinery, 2014, pp. 255–258.

[29] M. A. Tóth, M. L. G. Lecumberri, Y. Tang, and M. Cooke, "A corpus of noise-induced word misperceptions for Spanish," *J. Acoust. Soc. Am.*, vol. 137, no. 2, pp. EL184–EL189, 2015.

[30] H. Kawahara, I. Masuda-Katsuse, and A. de Cheveigné, "Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," *Speech Communication*, vol. 27, no. 3–4, pp. 187–207, 1999.

[31] V. Aubanel, M. L. G. Lecumberri, and M. Cooke, "The Sharvard Corpus: A phonemically-balanced Spanish sentence resource for audiology," *International Journal of Audiology*, vol. 53, no. 9, pp. 633–638, 2014.

[32] K. O. McGraw and S. P. Wong, "Forming inferences about some intraclass correlation coefficients," *Psychological Methods*, vol. 1, no. 1, pp. 30–46, 1996.

[33] D. V. Cicchetti, "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology," *Psychological Assessment*, vol. 6, no. 4, pp. 284–290, 1994.

[34] J. Z. Bakdash and L. R. Marusich, "Repeated Measures Correlation," *Frontiers in Psychology*, vol. 8, p. 456, 2017.

[35] V. C. Tartter, H. Gomes, and E. Litwin, "Some acoustic effects of listening to noise on speech production," *J. Acoust. Soc. Am.*, vol. 94, no. 4, pp. 2437–2440, 1993.

[36] I. Weiner, A. Healy, and R. Proctor, *Handbook of Psychology, Experimental Psychology*. Wiley, 2012, ch. Sensory processes – Audition.

[37] S. Gordon-Salant and P. J. Fitzgibbons, "Effects of stimulus and noise rate variability on speech perception by younger and older adults," *J. Acoust. Soc. Am.*, vol. 115, no. 4, pp. 1808–1817, 2004.

[38] M. Cooke and Y. Lu, "Spectral and temporal changes to speech produced in the presence of energetic and informational maskers," *J. Acoust. Soc. Am.*, vol. 128, no. 4, pp. 2059–2069, 2010.

[39] B. Larsby, M. Hällgren, B. Lyxell, and S. Arlinger, "Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects," *International Journal of Audiology*, vol. 44, no. 3, pp. 131–143, 2005.

[40] M. Krueger, M. Schulte, M. A. Zokoll, K. C. Wagener, M. Meis, T. Brand, and I. Holube, "Relation Between Listening Effort and Speech Intelligibility in Noise," *American Journal of Audiology*, vol. 26, no. 3S, pp. 378–392, 2017.

[41] Koopmans-van Beinum, Florien J., "Spectro-temporal reduction and expansion in spontaneous speech and read text: focus words versus non-focus words," in *Phonetics and Phonology of Speaking Styles*, 1991.