



Talker intelligibility and listening effort with temporally modified speech

Maximillian Paulus, Valerie Hazan, Patti Adank

Speech, Hearing and Phonetic Sciences, University College London, UK

m.paulus@ucl.ac.uk, v.hazan@ucl.ac.uk, p.adank@ucl.ac.uk

Abstract

Individual differences in talker acoustics substantially affect intelligibility in adverse listening conditions. Spectral enhancement has been found to reliably boost intelligibility in noise while temporal enhancement remains less effective. A potentially mediating factor that has been ignored so far is listening effort, as objectively assessed by the pupil dilation response. In two perception experiments, we measured intelligibility (keyword recall scores) and listening effort (pupil dilation) for two talkers in two listening conditions and with varying degrees of temporal modification. Results suggest that while keyword recall scores are sensitive to individual talker differences across listening conditions, the pupil dilation response reflects the degree of temporal and spectral distortion introduced by the signal processing techniques.

Index terms: listening effort, pupillometry, speech intelligibility, temporal modifications

1. Introduction

Talkers differ in their degree of intelligibility in adverse listening conditions [1]. Furthermore, large-scale studies with multiple listening conditions suggest that talker intelligibility is preserved across such conditions [2]. Traditionally, talker intelligibility is quantified perceptually by measuring keyword recall in noise. This approach is not sensitive to the effort experienced during listening, an increasingly important aspect of research into speech perception in noise [3], which can be measured by recording the pupil dilation. It has been shown that the pupil dilation increases with decreasing intelligibility in noise [4, 5, 6]. However, there is also evidence that at fixed intelligibility levels, pupil dilation is sensitive to the listening condition [7, 8].

Relevant to research into listening effort are foremost the characteristics of the listener such as nativeness or hearing status [6, 9]. However, it has been argued that talker-related characteristics such as regional or foreign accent should be considered within the proposed framework of listening effort [10]. A recent study [11] found decreased pupil dilation for speech produced in noise (Lombard speech [12]) when compared to speech produced in quiet. It has to be noted that at the tested SNRs, intelligibility was also increased for Lombard speech which corresponds to the established relationship of pupil dilation and intelligibility [6].

In the current study, we were specifically interested in the contribution of temporal features of speech to intelligibility and listening effort. The literature on algorithmic speech modifications suggests an asynchronous effect of manipulating speaking rate. While decreasing speaking rate does not improve intelligibility in noise [13, 14], increasing it impairs intelligibility when the compression factor is sufficiently large [15] which has been linked to the disruption of neural entrainment to speech [16, 17]. Listeners can only benefit from decreased speaking rate in the presence of a fluctuating masker when temporal dips can be exploited [18]. In natural speech, some studies find speaking rate

to be correlated with intelligibility [1], others do not [19], which is possibly mediated by the selection of stimuli or the task instructions. Our hypothesis was that a reduced speaking rate, artificial or natural, causes a decrease in listening effort, even when this is not reflected in recall scores per se. This finding could be potentially relevant for the design of applications employing user-driven speaking rate adjustments.

Experiment 1 aimed at measuring intelligibility and listening effort in two listening conditions for two talkers with distinct speaking rate. We also tested local and global temporal modifications of both talkers. Experiment 2 employed an adaptive procedure to assess listening effort at a fixed intelligibility level for a subset of conditions from experiment 1.

2. Methods

2.1. Participants

In experiment 1, sixteen normal-hearing native speakers of British English were recruited (12 females; $M_{Age} = 21.94$; age range, 18-29). In experiment 2, twenty-four normal-hearing native speakers of British English were recruited (12 females; $M_{Age} = 23.54$; age range, 19-33). They were either reimbursed for their participation or credits were assigned. Normal hearing ability was established by pure-tone audiometry at the beginning of each testing session. Thresholds for eligibility were 25 dB HL in both ears.

2.2. Materials

We used 192 IEEE sentences [20] spoken by two male speakers of Southern British English (Age: [71,69]) with intelligibility and speaking rate differences. Intelligibility was assessed in a pilot experiment with sixteen listeners. The talkers were recorded anechoically by the first author as part of a larger corpus. All sentences were preprocessed using Matlab [21] by converting to mono, down sampling to 22.5 kHz and applying a Butterworth high pass filter at 50 Hz. Before and after modifications were applied, the amplitudes of the audio files were root-mean-square (RMS) normalised. Speaking rates were 3.52 syllables/s for the intelligible talker (T08) and 4.33 syllables/s for the less intelligible talker (T13).

2.2.1. Local and global modifications

In addition to the unmodified speech (Plain), we created locally aligned versions of both talkers (Align), using the software VoCAAlign [22] similar to Cooke et al. [14]. Global modification was applied using the WSOLA algorithm [23]. We increased speaking rate by a factor 2 (Fast) and reduced it by a factor .66 (Slow). Overall sentence duration decreased by 50% (Fast) and increased by ~52% (Slow). All items were manually checked for signal processing artefacts.

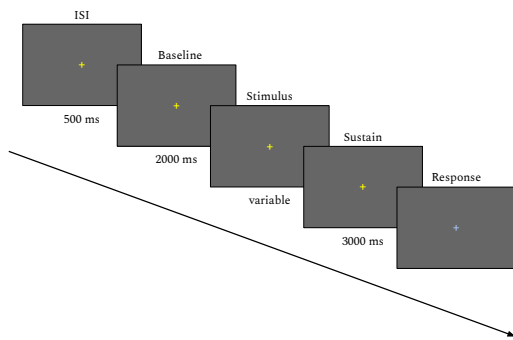


Figure 1: *Timeline of a trial in experiment 1.*

2.2.2. Listening conditions

In experiment 1, speech-modulated noise (SMN) and a noise-vocoder (Nx) were used to mask or distort the speech material. In experiment 2, only SMN was used. We preferred speech-modulated noise over competing speech in order to avoid effects of informational masking, which are known to affect pupil dilation [7]. We generated a speech-shaped noise using the LTAS of another male talker. For modulation, a set of non-experimental IEEE sentences spoken by the same talker was chosen. They were concatenated after removing the first second of the beginning and end of each sentence, ensuring that no silent parts were included. The amplitude envelope was then extracted using the Hilbert transform and applied to the speech-shaped noise. We generated 12 different versions of the SMN masker.

The Nx condition consisted of a 10-channel noise-vocoder spanning 30-9000 Hz. The original signal was divided into 10 frequency bands spaced according to the cochlear frequency-position function [24]. Amplitude envelopes were extracted from each band by applying a 4th-order Butterworth low-pass filter with a cutoff frequency at 256 Hz and half-wave rectification. The envelopes were then used to modulate respective white noise bands.

2.3. Design and Procedure

2.3.1. Experiment 1

We constructed 16 lists (one per condition) with 12 sentences per list. The lists were optimised so that the mean duration and predictability from context was approximately matched across lists ($SD_{Duration} \leq 0.05s$, $SD_{Entropy} \leq 0.3bits$). Entropy measures were based on a language model of British English (BNC corpus). It was ensured that the same keyword was not repeated more than twice within the same list which turned out to be the strictest constraint possible given the set of sentences. The lists and sentences within each list were randomised for each participant. Listening conditions (SMN and Nx) were divided over the two halves of the experiment and the order was counterbalanced across participants. Talkers and modifications were also counterbalanced within each half, according to a Latin square design. One of the 12 versions of the SMN was randomly selected for each noise trial.

At the beginning of each testing session, participants' hearing ability was measured using pure-tone audiometry. Hearing thresholds were determined at 1, 2, 4, 8, 0.25 and 0.5 kHz, in the

order indicated. After written and oral instructions, participants were asked to put their head comfortably on a head rest. The distance from rest to screen was kept constant at 55cm. Participants wore headphones (Sennheiser HD 25 SP II) during the experiment. A fixation cross was present throughout the experiment at the centre of the screen. The experiment was implemented in Matlab [21].

Each experiment half started with either 8 (SMN) or 16 (Nx) practice trials. We used more practice trials for the Nx conditions because of possible adaptation effects (e.g., [25]). The practice materials were taken from a female talker not included in the main experiment. Practice sentences were not contained in the experimental lists. After an inter-stimulus interval (ISI) of 500 ms, the baseline pupil size was recorded for 2000 ms. In masking conditions, noise started with the onset of the baseline. After the sentence was played, the pupil size was tracked for another 3000 ms. Participants repeated back as many words as possible and the correctly identified content words were selected by the experimenter on a separate control screen. Incorrect suffixes such as plural (-s) or tense (-ed) endings were allowed. The light settings inside the room were kept at approximately 130 lux and were only adjusted if a participant's pupil size measured during practice was too large or too small to show a response. A timeline of an experimental trial of experiment 1 is shown in Figure 1.

2.3.2. Experiment 2

The design and procedure of Experiment 2 was subject to the following changes. A subset of modifications was used (Plain, Align) with only one listening condition (SMN). We determined SNRs adaptively by measuring speech reception thresholds (SRTs) at 50% intelligibility for each talker and modification. We applied a one-up-one-down procedure [26]. Based on pilot data, initial SNRs were determined as -10 dB SNR (T08) and -6 dB SNR (T13). The step size was 2 dB. Final SNRs were calculated as the mean SNR of SRT trials after the fifth reversal. Eight lists (one per condition and test type) with 24 sentences per list were used. Eight trials with SMN masker were presented as practice at the beginning of the experiment. Modifications were distributed over the two halves of the experiment. The order was counterbalanced across participants, so was the order of talkers within each half. SRTs were always measured at the beginning of each half. The inter-stimulus interval was increased to 2000 ms in order to leave more time for the pupil size to return to baseline. Noise started after 1000 ms.

2.4. Pupil data preprocessing

Gaps of missing pupil data (blinks) were extended to 80 ms before and 140 ms after the gap. This procedure was suggested by Zekveld et al. [27] because eye-lid movements can affect pupil size even before or after complete closure. Trials that contained more than 40% missing data were removed from the analysis. Gaps were interpolated linearly. Outliers were defined as values outside 1.5 interquartile ranges below the first quartile or above the third quartile (Tukey's fences, [28]). They were replaced by the value of the 95th and the 5th percentile, respectively. For each trial, the mean of the second half of the baseline was calculated and the percentage change measured for each data point in that trial. A single value (peak dilation) was detected by finding the maximum dilation within a range from 1000 ms after sentence onset and 3000 ms after sentence offset.

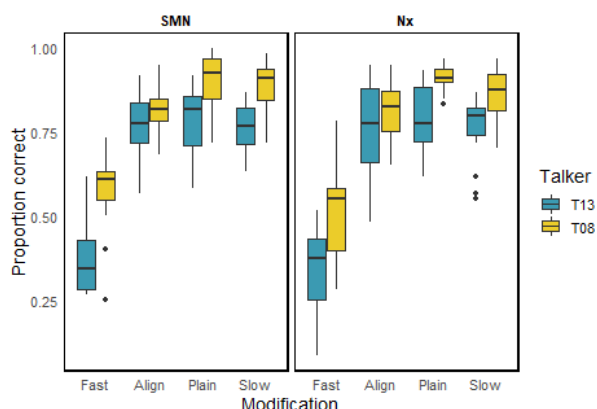


Figure 2: Proportion of keywords recalled correctly for all conditions in experiment 1.

2.5. Analysis

The proportion of content words (keywords) correctly identified and the peak pupil dilation were averaged across trials for each condition and listener. Both keyword recall and pupil dilation data were analysed using linear mixed effects models [29] with Satterthwaite’s degrees of freedom method. In all models, we allowed random intercepts for listeners. For multiple comparisons, Bonferroni adjustments were made.

3. Results

3.1. Behavioural

3.1.1. Experiment 1

Here we observed a main effect of talker [$F(1, 225) = 133.46, p < .001$], a main effect of modification [$F(3, 225) = 347.2, p < .001$] and an interaction of talker and modification [$F(3, 225) = 8.04, p < .001$]. Pairwise comparisons showed that T08 was more intelligible than T13 [$t = 11.55, p < .001$]. Fast speech was less intelligible than any other modification [$p < .001$] and aligned speech was less intelligible than plain speech [$t = -3.71, p < .05$]. The interaction showed that for T08 aligned speech was less intelligible than plain speech [$t = -4.29, p < .001$], while for T13 aligned speech was not more or less intelligible than plain speech [$p > .05$]. Figure 2 shows the proportion of keywords recalled correctly for all conditions in experiment 1. The full set of contrasts in the interaction can be retrieved at <https://goo.gl/FYUaXK>.

3.1.2. Experiment 2

There was no main effect or interaction of talker and modification on the proportion of keywords recalled correctly. This confirmed that the adaptive procedure was correctly applied. For SNRs, we found a main effect of talker [$F(1, 69) = 154.93, p < .001$] and modification [$F(1, 69) = 13.76, p < .001$], as well as an interaction of talker and modification [$F(1, 69) = 11.9, p < .001$]. Pairwise comparisons showed a lower SNR for T08 compared to T13 [$t = -12.45, p < .001$] and a higher SNR for aligned compared to plain speech [$t = 3.71, p < .001$]. The interaction showed a higher SNR for aligned speech of T08 [$t = 5.06, p < 0.001$] with no significant difference between modifications for T13 [$p > 0.05$]. Figure 3 shows SNRs for all conditions in experiment 2.

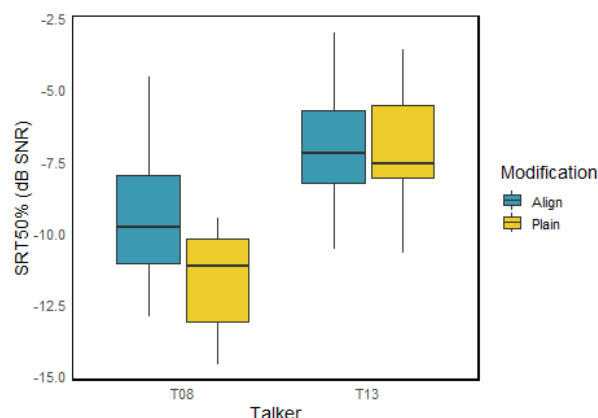


Figure 3: SNRs for all conditions in experiment 2.

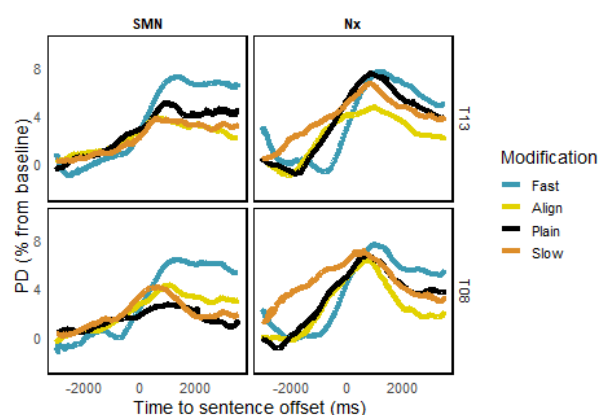


Figure 4: Pupil dilation (PD) from baseline for all conditions in experiment 1. Averaged across trials and listeners.

3.2. Pupillometry

For pupil data, all analyses are based on the peak pupil dilation, averaged across all trials of one condition.

3.2.1. Experiment 1

The only significant main effect we found was of condition [$F(1, 220.1) = 8.87, p < .05$] with larger peak pupil dilation for Nx. This effect is indicated in the time series depicted in Figure 4. In addition, the time series shows descriptive trends. They indicate that for Nx, most modifications have a large peak pupil dilation response whereas in the SMN condition responses were more distinct. For instance, fast speech showed an increased response for both talkers. For T13, aligned and slow speech show reduced peaks over plain whereas for T08 this trend is reversed.

3.2.2. Experiment 2

We found a main effect of modification [$F(1, 65.04) = 10.36, p < .05$], but no interaction with talker, contrary to the keyword recall data. Aligned speech of both talkers was associated with a larger peak pupil dilation, as indicated by the time series in Figure 5.

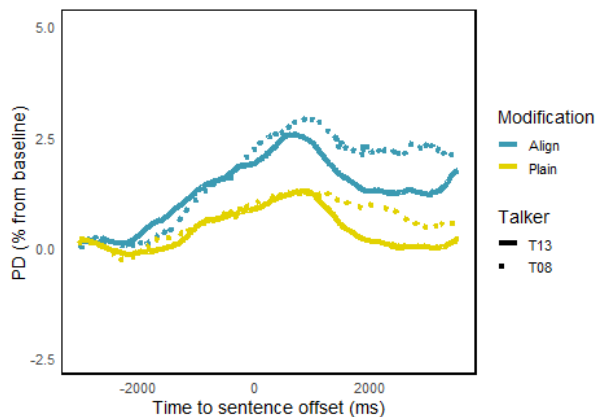


Figure 5: Pupil dilation (PD) from baseline for all conditions in experiment 2. Averaged across trials and listeners.

4. Discussion

Our first experiment was designed to measure intelligibility and listening effort with respect to two talkers varying in their speaking rate. We artificially manipulated speaking rate locally and globally and presented sentences in two listening conditions. Our aim was to assess the effect of temporal modifications on intelligibility and listening effort. Experiment 2 investigated only a subset of the conditions which allowed us to increase the number of tested items. Furthermore, we determined a fixed intelligibility level with overall lower SNRs, enabling an interpretation of pupil dilation independent of intelligibility.

We found that talker T08 was more intelligible overall, independent of the listening condition. This result shows that talker intelligibility can be preserved across listening conditions [2]. We believe that this result is due to a combination of acoustic features, which boost intelligibility selectively in certain conditions [1]. In both experiments, artificially increasing speaking rate reduced intelligibility. Unlike Cooke and Aubanel [18], we did not observe a benefit of artificially decreasing the speaking rate, despite the use of a fluctuating masker. In experiment 1, intelligibility might have been overall too high to detect this benefit: Cooke and Aubanel [18] had an average plain speech intelligibility of 34% in the SMN condition. However, even when keeping intelligibility at 50% (experiment 2) we did not observe a benefit of decreasing speaking rate locally. An explanation that Cooke and Aubanel provide for their improvements is that simply elongating the sentences increases the amount of spectral glimpses [18]. Aligning T13 to T08 might simply not have provided a sufficient increase in the glimpsing proportion.

In experiment 1, pupil dilation showed no difference between talkers and modifications. This might have been due to a small amount of trials per condition (12) and a small number of participants (16). Descriptive results do indicate increased pupil dilation for fast speech and a reduction for aligned and slow speech for the less intelligible talker in the SMN condition. If significant, these results would have supported our hypothesis, in that a reduction of speaking rate in the context of a fast talker decreases listening effort. In experiment 1, pupil dilation was however sensitive to the listening condition manipulation. This effect was not observed with keyword recall scores only, indicating that the pupil dilation responds to a distinct process. We hypothesise that the added level of distortion introduced by the vocoder had an impact on the pupil dilation. A previous

study found that the pupil dilation response is correlated with the number of vocoder channels, including conditions at ceiling intelligibility [30]. Our results contradict a previous study by Zekveld et al. [31] who found a larger pupil dilation for SMN over Nx. However, the two studies are not directly comparable since the number of channels were adjusted for each participant in the study by Zekveld et al. [31].

Pupillometry results of experiment 2 show that at fixed intelligibility levels, artificially aligning speech, i.e., increasing (T08) and decreasing (T13) speaking rate, is associated with a larger peak pupil dilation. At the same time, SNRs indicated a difference between plain and aligned speech only for T08. This result could indicate that at fixed intelligibility levels, pupil dilation is sensitive to the artificial re-timing of the original speech. It has been shown previously that pupil dilation and subjective ratings can capture differences between synthetic and natural speech, even when intelligibility is optimal [32]. It is possible that the local re-timing process introduced temporal distortions, affecting certain acoustic parameters relevant for speech intelligibility such as amplitude modulations [33]. However, since intelligibility was preserved, the effect might be related to more cognitive processes. With regard to neural models of speech processing, modifying speaking rate locally might have disrupted neural entrainment [16, 17].

5. Conclusion

We presented results of two experiments investigating intelligibility and listening effort associated with two talkers in two listening conditions and with a range of temporal modifications. Talker intelligibility was preserved across listening conditions. Artificially increasing speaking rate reduced intelligibility while decreasing it did not provide any benefit, contrary to our predictions. Listening effort, however, as measured by the pupil dilation response, was larger for both, an artificial local increase and decrease of speaking rate. This result, together with a larger response found for vocoded speech, could indicate that the pupil dilation is sensitive to the amount of spectral and temporal distortion introduced into the speech signal. Further experiments should be devised to parametrically assess the differential effect of temporal modifications on intelligibility and listening effort. So far, we have only considered a small range of temporal manipulations and SNRs. These findings could be useful for the design of audiological tests and software applications which allow users to control the rate of speech.

6. Acknowledgements

The authors would like to thank Martin Cooke for his help with speech modifications and Anita Wagner for support with the pupillometry analysis. The first author has received funding from the EU’s H2020 research and innovation programme under the MSCA GA 675324 (ENRICH: www.enrich-ctn.eu).

7. References

- [1] V. Hazan and D. Markham, “Acoustic-phonetic correlates of talker intelligibility for adults and children,” *Journal of the Acoustical Society of America*, vol. 116, no. 5, pp. 3108–3118, 2004.
- [2] T. Bent, A. Buchwald, and D. B. Pisoni, “Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech,” *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2660–2669, 2009.
- [3] M. K. Pichora-Fuller, S. E. Kramer, M. A. Eckert, B. Edwards, B. W. Hornsby, L. E. Humes, U. Lemke, T. Lunner, M. Matthen,

- C. L. Mackersie, G. Naylor, N. A. Phillips, M. Richter, M. Rudner, M. S. Sommers, K. L. Tremblay, and A. Wingfield, "Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL)," *Ear and Hearing*, vol. 37, pp. 5S–27S, 2016.
- [4] S. E. Kramer, T. S. Kapteyn, J. M. Festen, and D. J. Kuik, "Assessing Aspects of Auditory Handicap by Means of Pupil Dilatation," *Audiology*, vol. 36, no. 3, pp. 155–164, 1997.
- [5] A. A. Zekveld and S. E. Kramer, "Cognitive processing load across a wide range of listening conditions: Insights from pupillometry," *Psychophysiology*, vol. 51, no. 3, pp. 277–284, 2014.
- [6] B. Ohlenforst, A. A. Zekveld, T. Lunner, D. Wendt, G. Naylor, Y. Wang, N. J. Versfeld, and S. E. Kramer, "Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation," *Hearing Research*, vol. 351, pp. 68–79, 2017.
- [7] T. Koelewijn, A. A. Zekveld, J. M. Festen, and S. E. Kramer, "Pupil dilation uncovers extra listening effort in the presence of a single-talker masker," *Ear and Hearing*, vol. 33, no. 2, pp. 291–300, 2012.
- [8] A. A. Zekveld, D. J. Heslenfeld, I. S. Johnsrude, N. J. Versfeld, and S. E. Kramer, "The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load," *NeuroImage*, vol. 101, pp. 76–86, 2014.
- [9] G. Borghini and V. Hazan, "Listening effort during sentence processing is increased for non-native listeners: A pupillometry study," *Frontiers in Neuroscience*, vol. 12, pp. 1–13, 2018.
- [10] K. J. Van Engen and J. E. Peelle, "Listening effort and accented speech," *Frontiers in Human Neuroscience*, vol. 8, pp. 1–4, 2014.
- [11] O. Simantiraki, M. Cooke, and S. King, "Impact of different speech types on listening effort," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2018, pp. 2267–2271.
- [12] E. Lombard, "Le signe de l'elevation de la voix," *Ann. Mal. de L'Oreille et du Larynx*, pp. 101–119, 1911.
- [13] V. Aubanel and M. Cooke, "Information-preserving temporal reallocation of speech in the presence of fluctuating maskers," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2013, pp. 3592–3596.
- [14] M. Cooke, C. Mayo, and J. Villegas, "The contribution of durational and spectral changes to the Lombard speech intelligibility benefit." *The Journal of the Acoustical Society of America*, vol. 135, no. 2, pp. 874–883, 2014.
- [15] E. Dupoux and K. Green, "Perceptual adjustment to highly compressed speech: effects of talker and rate changes." *Journal of experimental psychology. Human perception and performance*, vol. 23, no. 3, pp. 914–927, 1997.
- [16] K. V. Nourski, R. A. Reale, H. Oya, H. Kawasaki, C. K. Kovach, H. Chen, M. A. Howard, and J. F. Brugge, "Temporal Envelope of Time-Compressed Speech Represented in the Human Auditory Cortex," *Journal of Neuroscience*, vol. 29, no. 49, pp. 15 564–15 574, 2009.
- [17] A. L. Giraud and D. Poeppel, "Cortical oscillations and speech processing: Emerging computational principles and operations," *Nature Neuroscience*, vol. 15, no. 4, pp. 511–517, 2012.
- [18] M. Cooke and V. Aubanel, "Effects of linear and nonlinear speech rate changes on speech intelligibility in stationary and fluctuating maskers," *The Journal of the Acoustical Society of America*, vol. 141, no. 6, pp. 4126–4135, 2017.
- [19] A. R. Bradlow, G. M. Torretta, and D. B. Pisoni, "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics." *Speech communication*, vol. 20, no. 3, pp. 255–272, 1996.
- [20] Institute of Electrical and Electronics Engineers, "IEEE recommended practices for speech quality measurements," *IEEE Trans. Aud. Electroacoust.*, vol. 17, pp. 227–246, 1969.
- [21] MATLAB, *version 9.0.0 (R2016a)*. Natick, Massachusetts: The MathWorks Inc., 2016.
- [22] Synchroarts, "VocALign Project," 2018. [Online]. Available: www.synchroarts.com
- [23] M. Demol, W. Verhelst, K. Struyve, and P. Verhoeve, "Efficient non-uniform time-scaling of speech with WSOLA," in *Int. Conf. on Speech and Computers (SPECOM)*, 2005, pp. 163–166.
- [24] D. D. Greenwood, "A cochlear frequency position function for several species - 29 years later," *The Journal of the Acoustical Society of America*, vol. 87, no. 6, pp. 2592–2605, 1990.
- [25] J. E. Peelle and A. Wingfield, "Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 31, no. 6, pp. 1315–1330, 2005.
- [26] R. Plomp and A. M. Mimpen, "Improving the reliability of testing the speech reception threshold for sentences," *Audiology*, vol. 18, no. 1, pp. 43–52, 1979.
- [27] A. A. Zekveld, S. E. Kramer, and J. M. Festen, "Pupil Response as an Indication of Effortful Listening: The Influence of Sentence Intelligibility," *Ear and Hearing*, vol. 31, no. 4, pp. 480–490, 2010.
- [28] J. W. Tukey, *Exploratory data analysis (Vol. 2)*. Addison-Wesley, 1977.
- [29] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
- [30] M. B. Winn, J. R. Edwards, and R. Y. Litovsky, "The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation," *Ear and Hearing*, vol. 36, no. 4, pp. e153–e165, 2015.
- [31] A. A. Zekveld, M. Rudner, S. E. Kramer, J. Lyzenga, and J. Rönnerberg, "Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech," *Frontiers in Neuroscience*, vol. 8, pp. 1–11, 2014.
- [32] A. Govender and S. King, "Using pupillometry to measure the cognitive load of synthetic speech," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2018, pp. 2838–2842.
- [33] J. C. Krause and L. D. Braida, "Acoustic properties of naturally produced clear speech at normal speaking rates," *The Journal of the Acoustical Society of America*, vol. 115, no. 1, pp. 362–378, 2004.