PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

# Talker intelligibility and listening effort: the role of speaking rate

Maximillian PAULUS[1]; Valerie HAZAN[2]; Anita WAGNER[3]; Patti ADANK[4]

[1, 2, 4] University College London, UK

[3] Saarland University, Germany

## ABSTRACT

Listening to speech in noise induces increased cognitive effort and fatigue. Effort is mediated by listener characteristics such as nativeness or hearing status. Moreover, noise-related issues can be exacerbated by talker differences. Nevertheless, past research has so far not related talker differences to cognitive effort. Based on an anechoically recorded corpus of sentences spoken by sixteen Southern British English speakers, we conducted listening experiments in combination with pupillometry. We presented temporally or spectrally distorted speech (using time-compression and noise-vocoding, respectively). Furthermore, undistorted speech was presented in quiet and embedded in speech-shaped masking noise. Intelligibility measures were obtained based on keywords recalled correctly. Listening effort was measured by tracking pupil size over time. Our results show that peak pupil dilation increased for both distorted and masked speech. Peak pupil dilation also increased for less intelligible talkers in the temporally degraded condition. The results indicate a link between speaking rate differences and listening effort and will be discussed in the context of models of listening effort.

Keywords: listening effort, pupillometry, speech intelligibility, speaking rate

## 1. INTRODUCTION

Pupillometry is widely used to measure listening effort in normal-hearing or hearing-impaired populations (1). It has recently been suggested that talker-specific variations such as accent should be considered in frameworks of listening effort (2). Besides accents, talkers differ in their intrinsic intelligibility that can be determined by a combination of acoustic-phonetic features, including speaking rate (3). Intrinsic intelligibility is often measured in masking noise, but has been found to be preserved in other listening conditions such as those involving spectral degradations (4). Even though slower speaking rates have been observed for highly-intelligible talkers (3), other studies did not find consistent evidence for the role of speaking rate (5). The role of speaking rate has been investigated further by artificially shortening or lengthening utterances, with most studies concluding that speaking rate changes do not benefit intelligibility (6, 7). It is possible that a slower speaking rate, even in the absence of intelligibility benefits, bears a cognitive processing advantage. In recent years, pupillometry has been used to investigate the cognitive effects of listening to speech in noise (8). A recent study (9) investigated the effects of speaking rate on task-evoked pupil dilation when listening to speech in quiet. Across listener age-groups, no effect was found. This null result could have been due to the relative ease of adapting to faster speaking rates in noise-free environments (10). In a noisy or degraded listening environment, these effects might be elevated. In this study, we tested the impact of masking, spectral and temporal degradation on intelligibility and listening effort. We also investigated intrinsic speaking rate differences between talkers for the same dependent measures.

[1] m.paulus@ucl.ac.uk
[2] v.hazan@ucl.ac.uk
[3] anita.wagner@uni-saarland.de
[4] p.adank@ucl.ac.uk

## 2. Methods

### 2.1 Participants

Sixty-four normal-hearing native speakers of British English were recruited for the experiment (40 females; mean age = 22.26; age range, 18 - 37). They were either reimbursed for their participation following the guidelines of the Division of Psychology and Language Sciences at UCL or given credits. Hearing ability was established by a standardised audiometric test at the beginning of each testing session. Exclusion criteria were thresholds above 30 dB HL at any of the tested frequencies (0.25, 0.5, 1, 2, 4, 8 kHz). One participant was excluded post-hoc despite good hearing thresholds because he could not identify any words in the spectral condition.

### 2.2 Materials

We used anechoic recordings of 16 Southern British English speakers: eight older adults (four females; mean age = 71.00; age range, 61 – 77) and eight younger adults (four females; mean age = 26.75; age range, 22 - 33). Sentence material were low-predictability and phonetically balanced Harvard sentences (11) adapted for British English. We presented 192 sentences in total, divided into four lists of 48 items each, matched for duration. The same keyword did not appear more than twice within the same list. Items were rms-normalised in Praat (12) and the output level was calibrated at 70 dB SPL. We created a temporal distortion by time-compressing speech at a rate of 37 %. We applied the pitch-synchronous overlap-add implementation in Praat (12). We created a spectral distortion by dividing the original signal into six frequency bands spaced according to the cochlear frequency-position function (13). 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 white noise. We created a noise masker by extracting the long-term average spectrum of a separate set of recorded sentences and generating noise with the same spectrum. The masker was added to the original signal at a signal-to-noise of -1 dB.

### 2.3 Design and Procedure

Items were presented to each listener in all three experimental conditions (Spectral, Noise, Temporal) and in quiet, with a different talker in each condition. Each talker was therefore heard by four listeners in each condition, i.e., by 16 listeners in total. Participants wore headphones (Sennheiser HD 25 SP II) throughout the experiment. Pupil recordings were obtained using an EyeLink 1000 table-mounted eye-tracker at a distance of 55 cm from a chin rest. A sampling rate of 500 Hz was used. The light level was kept constant at 130 lux. For participants with very large or very small resting state pupil sizes, the light level was individually adjusted (14). The experiment started with eight practice trials in which sentences in quiet were presented to the participants. Practice talker and sentences were not contained in the experimental lists. After hearing a sentence, participants repeated back the identified words and the correctly identified words were selected by the experimenter on a separate control screen. The experiment was implemented in Matlab (15).

### 2.4 Analysis

Time series of pupil size data were preprocessed in the following way: first, gaps of missing data were extended (16). Trials that contained more than 40% missing data were removed. This resulted in an average of 42.95 trials per condition entering the analysis. We applied linear interpolation to missing values. Outliers were defined as Tukey's fences (17) and were replaced by the value of the 95th and the 5th percentile, respectively. Divisive baseline correction was applied using the mean of a pre-trial baseline. One single value was extracted for further analysis, the peak dilation, by finding the maximum dilation within a range from 1000 ms after sentence onset and 3000 ms after sentence offset. Both keyword recall and peak pupil dilation data were then aggregated across trials for each listener and condition. Statistically, we investigated differences between listening conditions, with listeners as random effects. We used linear mixed effects models (18) with Satterthwaite's degrees of freedom method. In all models, we allowed random intercepts. For multiple comparisons, Bonferroni adjustments were made. To obtain a measure of speaking rate for each talker, we measured the sentence duration from the onset of the first word to the offset of the last word. The number of syllables was calculated for each sentence in text form using the package quanteda in R (19) and the speaking rate was then defined as the number of syllables over sentence duration.

## 3. Results

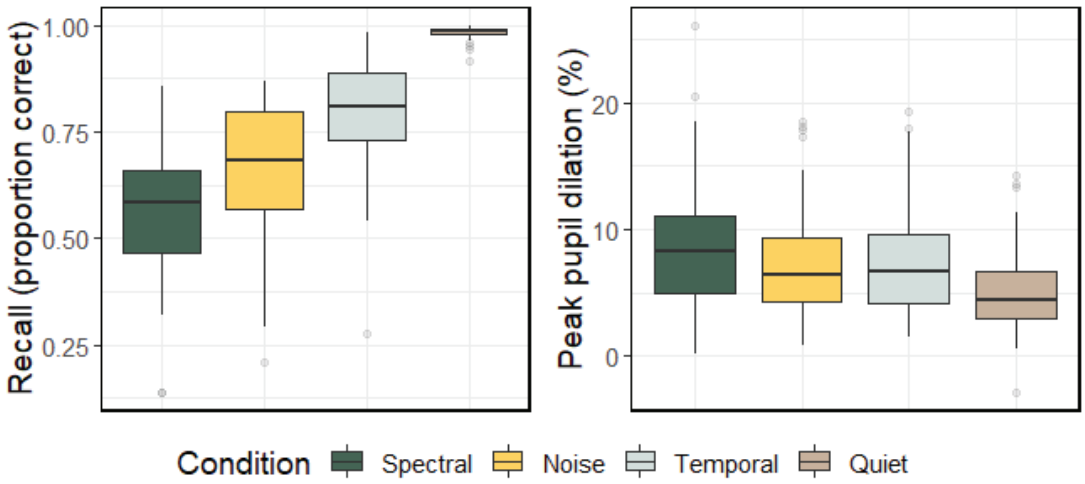


Figure 1 – Recall scores and peak pupil dilation for all listeners and conditions.

There was a significant main effect of condition for recall scores [$F(3,186) = 155.36, p < .001$]. Pairwise comparisons showed that recall was worse for all three manipulations compared to Quiet [$p < .001$]. Recall for Spectral was worse than recall for both Noise and Temporal [$p < .001$] while recall for Noise was worse than recall for Temporal [$p < .001$]. There was a significant main effect of condition for the peak pupil dilation [$F(3,178.7) = 28.71, p < .001$]. Pairwise comparisons showed that dilation was larger for all three manipulations compared to Quiet [$p < .001$]. Dilation for Spectral was larger than dilation for both Noise and Temporal [$p < .05$]. Dilation for Noise was not larger than dilation for Temporal, contrary to recall scores. The peak pupil dilation correlated negatively with recall scores only for Temporal [$r = -0.4, p < .05$].

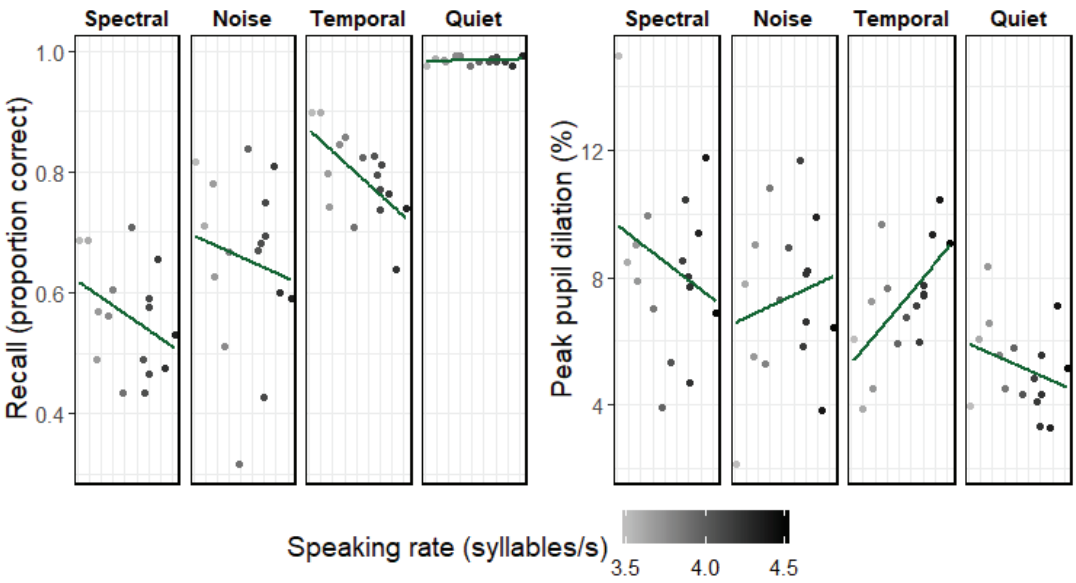


Figure 2 – Recall scores and peak pupil dilation as a function of talker speaking rate.

There was a significant negative correlation between speaking rate and recall scores for Temporal [$r = -0.63, p < .05$]. There was also a significant positive correlation between speaking rate and peak pupil dilation for Temporal [$r = 0.62, p < .05$]. Talkers with faster speaking rates were associated with lower recall scores, but larger peak pupil dilation. However, there was no significant correlation between peak pupil dilation and recall scores in any of the conditions.

## 4. Discussion

First, we investigated recall scores and peak pupil dilation for listeners in four conditions, in masking noise, with temporal and spectral distortion and in quiet. The peak pupil dilation was related to the severity of the distortion, which was reflected in recall scores. The relationship between intelligibility and peak pupil dilation has been found in previous studies (20). However, there was no significant difference in peak pupil dilation between masking and the temporal distortion condition, even though differences were observed for recall scores. As suggested previously, there is no linear relationship between pupil dilation and recall scores (14, 21). Effects on listening effort might therefore only have been pronounced when the intelligibility level differences were sufficiently large between conditions. On the other hand, correlation analyses showed a linear relationship between peak dilation and recall scores in the temporal distortion condition. This finding could be specific to the type of distortion or to the range of recall scores in that condition; the temporal distortion resulted in relatively high intelligibility levels (well above 50%) compared to the other two experimental conditions. It has been suggested recently that intelligibility levels around 50% (or lower) have to be treated with caution since disengagement from the task can be reflected in decreased pupil dilation (14). Our second investigation focused on the talker. We were specifically interested in speaking rate effects on recall scores and peak pupil dilation. In the temporal distortion condition, speaking rate was significantly correlated with both recall scores and peak pupil dilation. Talkers with faster speaking rate were associated with lower recall scores and larger peak pupil dilation. This result might indicate that intrinsic speaking rate differences per se do not cause increased listening effort, similar to a previous finding (9), but that artificial time-compression renders speaking rate a relevant factor for listening effort.

## 5. Conclusion

In the current study, we were able to relate talker-specific speaking rate differences to intelligibility and listening effort. Furthermore, listener-specific results could be linked to previous studies indicating the non-linear relationship between intelligibility and listening effort. A disadvantage of our study was that given the nature of the design, it was only possible to calculate recall and peak pupil dilation averages across small samples of listeners. Further studies will therefore be needed to systematically assess the impact of speaking rate on listening effort with and without temporal distortions.

## ACKNOWLEDGEMENTS

The authors would like to thank Outi Tuomainen for her help with acoustic analyses. The first author has received funding from the EU's H2020 research and innovation programme under the MSCA GA 675324 (ENRICH: www.enrich-etn.eu).

## REFERENCES

1. Kramer SE, Kapteyn TS, Festen JM, Kuik DJ. Assessing Aspects of Auditory Handicap by Means of Pupil Dilatation. Audiology. 1997;36(3):155-64.
2. Van Engen KJ, Peelle JE. Listening effort and accented speech. Frontiers in Human Neuroscience. 2014;8:1-4.
3. Hazan V, Markham D. Acoustic-phonetic correlates of talker intelligibility for adults and children. Journal of the Acoustical Society of America. 2004;116(5):3108-18.
4. Bent T, Buchwald A, Pisoni DB. Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech. The Journal of the Acoustical Society of America. 2009;126(5):2660-9.
5. Bradlow AR, Torretta GM, Pisoni DB. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. Speech communication. 1996;20(3):255-72.
6. Picheny MA, Watson TJ, Durlach NI, Braida LD. Speaking Clearly for the Hard of Hearing III. Journal of Speech and Hearing Research. 1989;32(September):600-3.
7. Uchanski RM, Choi SS, Braida LD, Reed CM, Durlach NI. Speaking clearly for the hard of hearing .4. Further studies of the role of speaking rate. Journal of Speech and Hearing Research. 1996;39(3):494-509.
8. Zekveld AA, Koelewijn T, Kramer SE. The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. Trends in Hearing. 2018;22:1-25.

9. Koch X, Janse E. Speech rate effects on the processing of conversational speech across the adult life span. The Journal of the Acoustical Society of America. 2016;139(4):1618-36.
10. Adank P, Janse E. Perceptual learning of time-compressed and natural fast speech. Journal of the Acoustical Society of America. 2009;126(5):2649-59.
11. Institute of E, Electronics E. IEEE recommended practices for speech quality measurements. IEEE Trans Aud Electroacoust. 1969;17:227-46.
12. Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program]. Version 6.0.40. 2018.
13. Greenwood DD. A cochlear frequency position function for several species - 29 years later. The Journal of the Acoustical Society of America. 1990;87(6):2592-605.
14. Wendt D, Koelewijn T, Książek P, Kramer SE, Lunner T. Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. Hearing Research. 2018;369:67-78.
15. Matlab. version 9.0.0 (R2016a). Natick, Massachusetts: The MathWorks Inc.; 2016.
16. Siegle GJ, Ichikawa N, Steinhauer S. Blink before and after you think: blinks occur prior to and following cognitive load indexed by pupillary responses. Psychophysiology. 2008;45(5):679-87.
17. Tukey JW. Exploratory data analysis (Vol. 2): Addison-Wesley; 1977.
18. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015;67(1):1-48.
19. Benoit K. quanteda: Quantitative Analysis of Textual Data. 2018.
20. Zekveld AA, Kramer SE, Festen JM. Pupil Response as an Indication of Effortful Listening: The Influence of Sentence Intelligibility. Ear and Hearing. 2010;31(4):480-90.
21. Ohlenforst B, Zekveld AA, Lunner T, Wendt D, Naylor G, Wang Y, et al. Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. Hearing Research. 2017;351:68-79.