Environmental noise affects audiovisual gain during speech comprehension in adverse listening conditions

The Communicative Brain
Noor Seijdel*[1], Sara Mazzini*[1] & Linda Drijvers*[1,2]
* all authors contributed equally

[1] Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[2] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands

Abstract
Face-to-face communication involves both auditory speech and visual bodily information (e.g. visible speech, gestures), and often occurs in noisy settings. In such settings, meaningful hand gestures have shown to facilitate speech comprehension. However, individual differences exist in how much listeners benefit from these gestures. One factor that could impact how a listener processes and integrates audiovisual information is their experience with environmental noise. For example, growing up or going to school in high-noise environments (such as in the proximity of frequent heavy traffic or airports) has been shown to influence language abilities, such as reading comprehension, and speech perception. In the current study, we investigated whether a listener's experience with environmental noise affects how much they can benefit from gestures in adverse listening conditions. In an online experiment, 40 participants watched video clips of an actress articulating an action verb, accompanied by an iconic gesture or none, in clear or degraded speech. Participants identified the verb and rated the noise levels of their current and previous living environments based on their geographical locations. We hypothesized that high-noise environments would hinder task performance and audiovisual gain. The results indicated that the amount of environmental noise, in current and previous living environments, influenced how much participants benefited from gestures during degraded speech comprehension. Individuals from lower-noise environments benefitted more from visual semantic information than those from high-noise environments. Individuals from higher-noise environments performed better overall, and benefitted less from gestures, potentially because they developed adaptive mechanisms to cope with auditory degradation.

1 Introduction

Face-to-face communication often takes place in noisy settings, and is a complex process that relies on processing and integrating auditory and visual information. Previous research has demonstrated that visual bodily cues, such as visible articulatory movements and hand gestures, can significantly enhance speech comprehension, particularly in adverse listening conditions where auditory information may be degraded (Drijvers & Özyürek, 2017). For example, iconic gestures, that depict or represent objects or actions, may provide additional semantic information that complements speech (McNeill, 1992), or makes it easier to understand a noisy speech signal. In line with this, previous research has shown that the semantic information from iconic gestures can impact speech comprehension, at both behavioral and neural levels (Drijvers et al., 2017, 2018; Holle et al., 2010) This integration of auditory and visual cues, known as audiovisual speech-gesture integration, is essential for effective communication in a variety of settings.

Despite the well-established role of iconic gestures in facilitating degraded speech comprehension, individual differences may exist in the extent to which listeners benefit from these cues. One factor that could contribute to these individual differences is the presence and amount of environmental noise in a listener's living environment. Environmental noise, defined as unwanted or harmful outdoor sounds created by human activities (European Directive on environmental noise, 2002), is a pervasive aspect of daily life for many individuals, with sources such as heavy traffic, construction work, and airports contributing to this auditory landscape.

Environmental noise exposure is thought to affect the brain and cognition due to its detrimental effects on learning (e.g., distracting attention), sleep, stress, and an individual's sense of control and self-efficacy (Thompson et al., 2022). Exposure to high levels of environmental noise has been shown to negatively impact various aspects of communication, including language abilities, such as reading comprehension, and unimodal speech perception (Evans & Maxwell, 1997; Maxwell & Evans, 2000; Hygge et al., 2002; Lercher et al., 2003; Seabi et al., 2012). The detrimental effects of noise on communication may be due, in part, to increased cognitive load, as individuals in noisy environments must allocate additional cognitive resources to process and filter out irrelevant auditory information.

Relevant to the current work, this increased cognitive load could potentially impair the ability to process language and integrate audiovisual information, leading to a reduced benefit from gestures during degraded speech comprehension. This would be in line with work showing that visual peripheral cues are detected less often in noise (Hockney, 1970). However, the relationship between environmental noise exposure and the benefit of gestures during degraded speech comprehension has not been investigated yet. The current study aims to fill this gap by examining whether exposure to environmental noise levels affects the degree to which listeners can benefit from gestures in degraded speech comprehension. Given the negative impact of environmental noise on unimodal language processing (Evans & Maxwell, 1997; Maxwell & Evans, 2000; Hygge et al., 2002; Lercher et al., 2003; Seabi et al., 2012), we hypothesized that individuals who live(d) in high-noise environments would perform worse on an audiovisual speech comprehension task and benefit less from gestures during degraded speech comprehension.

In an online experiment, participants were presented with video clips of an actress uttering a highly-frequent action verb accompanied by either an iconic gesture or no gesture, and the speech was presented in either clear or degraded (6-band vocoded) conditions. Participants' task was to identify the verb conveyed in the videos. Additionally, we collected information on participants' current and previous living environments' noise levels, based on a detailed atlas provided by the Dutch Institute for

Public Health and the Environment for sound disturbance (Rijksinstituut voor Volksgezondheid en Milieu, 2020).

The results of this study have important implications for understanding individual differences in audiovisual speech-gesture integration and the impact of environmental noise on multimodal speech comprehension in adverse listening conditions. By investigating how previous auditory experiences can affect audiovisual speech-in-noise comprehension and auditory processing, these findings may inform the design of interventions to enhance communication in noisy environments. Furthermore, this research may inspire future studies to explore the underlying neural mechanisms of these effects, the potential long-term consequences of noise exposure on audiovisual integration, and interventions that can facilitate audiovisual integration in individuals with neurodiverse traits.

2 Methods

2.1 Participants

160 participants native speakers of Dutch (116 females, 18-40 years old) took part in the online experiment. Participants were recruited through the Max Planck Institute for Psycholinguistics participant database, social media, and advertisements on forums. As part of a different research line, participants were assigned to one of four groups: 1) healthy controls, 2) ADHD, 3) Dyslexia, or 4) Autism. For the current study, 40 healthy controls were included (30 females, mean age = 24.1, SD = 4.17). All participants reported no hearing impairments and had normal or corrected-to-normal vision. All participants gave their informed consent preceding the experiment.

2.2 Stimuli

Videos were selected from a larger set of 240 videos used in previous work (see e.g., Drijvers & Ozyürek, 2017; Drijvers, Van der Plas, Ozyürek & Jensen, 2019; Drijvers, Jensen & Spaak, 2021). Participants were presented with 160 video clips showing an actress uttering a highly-frequent action verb accompanied by an iconic gesture or no gesture (for a detailed description of pretests on recognizability and iconicity of the gestures, see Drijvers & Ozyürek, 2017). Auditory information could be clear or degraded (6-band noise-vocoded, see Drijvers & Ozyurek, 2017). All gestures used in the videos were rated as potentially ambiguous when viewed without speech.

2.3 Experimental procedure

The experiment was conducted online using the Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2020). Participants were asked to wear headphones and the experiment started with a sound check, which allowed participants to set a comfortable volume and checked whether a headphone was used. After these checks, participants first completed a questionnaire, in which they indicated the environmental noise levels in their current and previous (“where they grew up”) living environments (see 2.4 for details).

During the experiment, participants performed a free-recall task. All trials started with a fixation cross (1000ms), followed by the video. Participants were asked to press the space bar as fast as possible when they successfully recognized which verb was communicated during the video. After pressing the spacebar, participants could type in the verb without any time pressure. The stimuli were presented in a randomized order and divided over blocks of 40 trials, with a self-paced rest after every block.

After the experiment, participants filled in a short questionnaire about their hearing and listening experiences and hearing and listening ability in different situations (15-item Speech, Spatial, and Qualities of Hearing scale questionnaire; self-reported hearing, see Moulin et al., 2019). Finally, we asked them about their experiences during the experiment (if there were any problems with the videos or if anything happened during the experiment).

2.4 Environmental noise levels
We asked participants to identify the environmental noise levels of their current and previous living environments, by using the atlas from the Dutch Institute for Public Health and the Environment for sound disturbance (Rijksinstituut voor Volksgezondheid en Milieu, 2020). Data from the following noise sources were used to calculate noise levels: state highways, municipal roads and provincial roads (data from 2019-2021), rail traffic (data from 2019), aviation (data from 2016), industry (numerical estimate from 2008) and wind turbines (data from 2021). A separate noise map was first created for each of these sources. These were then merged into one map, giving a global picture of the amount of noise at a given location. To express the amount of noise, Lden (Level Day-Evening-Night) was used. This is a measure for calculating the average amount of noise per 24-hour period. The categories from 'very good' to 'very bad' come from the Explanatory Memorandum to the Noise Abatement Regulations. This classification is derived from studies into the relationship between noise and annoyance. All noise sources count equally (no weighting was done). The accumulated noise level is the summation of the noise of the various noise source types, expressed in decibel (dB).

2.5 Data analysis
We calculated the Token Sort Ratio for all responses. Token Sort Ratio is a form of fuzzy string matching that calculates the similarity of two strings (e.g., the correct answer and the participant's response) and returns a percentage overlap between 0 and 100 (see https://github.com/seatgeek/fuzzywuzzy). We have excluded responses that were slower than 4000 ms and faster than 300 ms (5.55% of the data, 355 entries). These responses often included solely a question mark, or in some cases a short sentence that stated that the participant did understand the speech in the video.

As a first step, we established whether we could replicate the gestural enhancement effect for degraded speech comprehension (as observed in e.g., Drijvers & Ozyurek, 2017; Wilms et al., 2022). For this purpose, we fitted a linear mixed effects model using lme4 (version; Bates et al., 2015) in R (4.2.3, R Core Team, 2023), and obtained P-values with the lmerTest (3.1.1, Kuznetsova et al., 2017) package. We used a maximal random effects structure. Both fixed factors (Speech/Gesture) were sum-to-zero contrast coded (-0.5, 0.5). When the model did not converge with the full random effects structure, we used the estimates of this non-converged model fit as starting values to restart the fit, and compared the estimates from different optimizers using the allFit() function. If these yielded similar values, the warnings were considered as false positives. This was the case for all reported models.

As a second step, we then tested whether the level of noise a listener experienced in their previous and current living environment affected this gestural enhancement effect. Here we used a data-driven model-building procedure, starting with a model that included the interaction between speech and gesture, and other factors step-wise. We did this separately for the effect of the current and previous living environment, and only kept interactions in the model when they significantly improved the model fit.

3 Results

3.1 Gestures enhance degraded speech comprehension
In line with previous work (e.g., Drijvers & Ozyürek, 2017), we observed a clear gestural enhancement effect of degraded speech comprehension. Participants' comprehension of the degraded speech signal was significantly better when a gesture was present, and this gestural enhancement was larger when speech was degraded than when speech was clear (β = 17.38, *SE* = 4.06, t(105.25)= 4.26, *p* = <.001, clear speech: mean TSR = 99.00%, SD = 6.39% clear speech + gesture: mean TSR = 97.83%, SD = 11.40%; degraded speech: mean TSR = 64.17%, SD = 33.65%, degraded speech + gesture: mean TSR = 80.63%, SD = 29.95).

3.2 The noise level of current and previous living environment determines how much a listener benefits from gestures during degraded speech comprehension

A model containing a three-way interaction between Speech, Gesture and Current Noise Environment was a significantly better fit than a model not containing Current Noise Environment ($\chi^2(40)=217.89$, $p<.001$). Similarly, a model containing a three-way interaction between Speech, Gesture and Previous Noise Environment was a significantly better fit than a model without Previous Noise Environment ($\chi^2(40)=98.09$, $p<.001$). Both analyses revealed that listeners who live(d) in a quiet environment, benefit more from gestures when speech is degraded than listeners who live(d) in a noisy environment (see Figure 1 and 2).

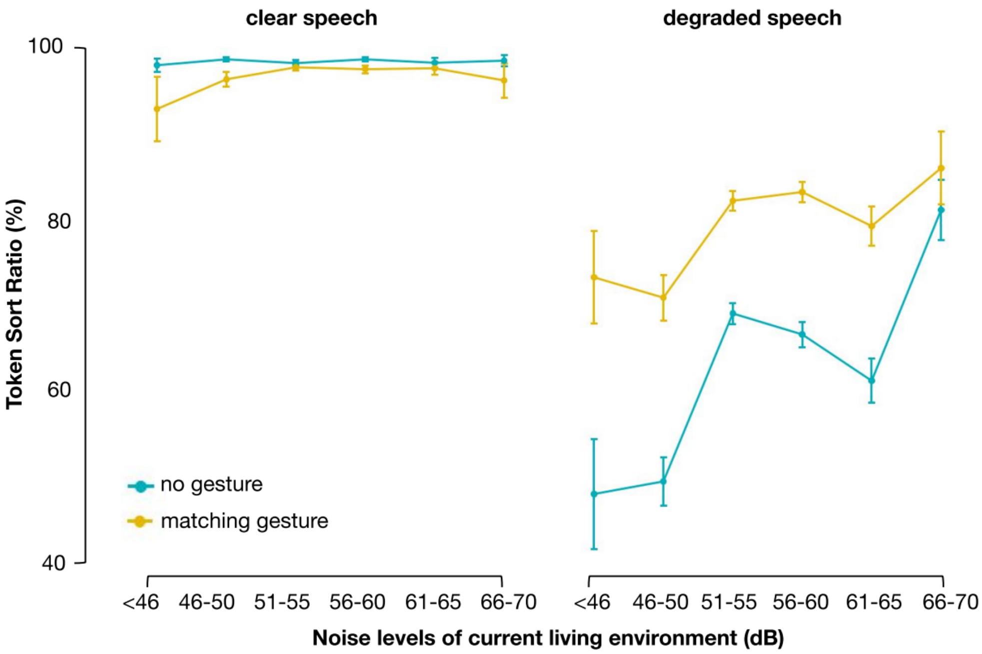

Figure 1 Gestures enhance degraded speech comprehension more for participants living in a quiet environment. Token Sort Ratio (%) per condition. Error bars represent SE.

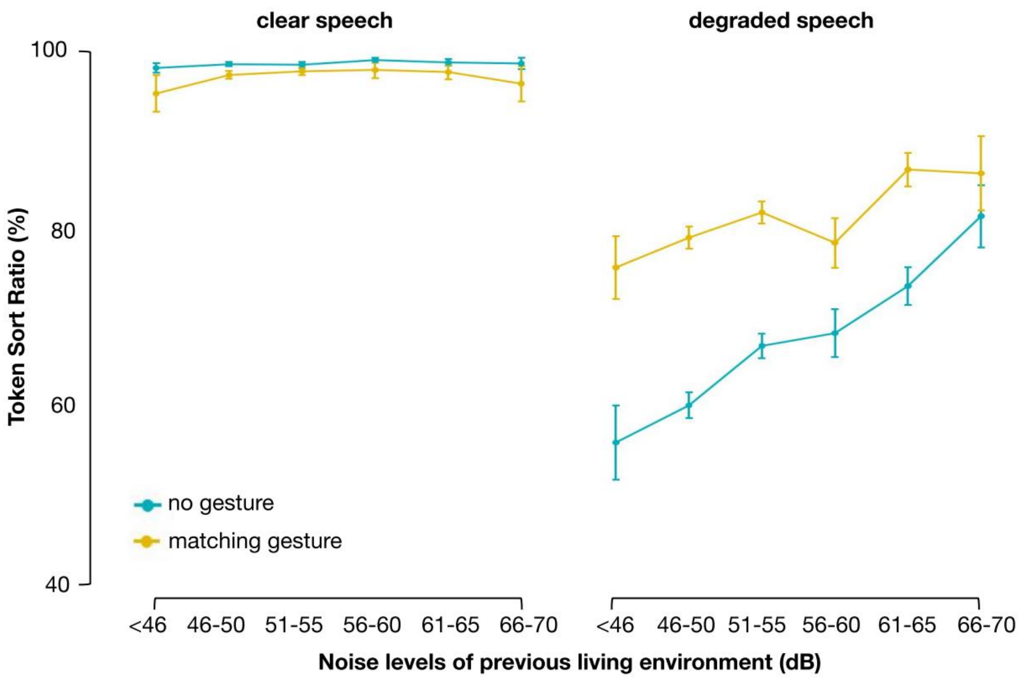

Figure 2 Gestures enhance degraded speech comprehension more for participants who grew up in a quiet environment. Token Sort Ratio (%) per condition. Error bars represent SE.

4 Discussion

The present study investigated the impact of environmental noise on the ability of listeners to benefit from visual semantic information conveyed by hand gestures during degraded speech comprehension. Our results showed that environmental noise exposure in current and previous living environments influenced the degree to which individuals benefit from gestures in degraded speech comprehension.

Gestural enhancement of degraded speech comprehension is larger for individuals who experienced lower levels of environmental noise

Consistent with our hypotheses, individuals who experienced lower levels of environmental noise in their living environments exhibited a larger benefit from gestures during degraded speech comprehension. These results align with previous research highlighting the positive role of visual cues in facilitating comprehension under challenging listening conditions (Holle et al., 2010; Drijvers et al., 2017;2018). Moreover, our findings suggest that reduced exposure to environmental noise enhances the integration of visual cues, such as gestures, in the context of degraded speech comprehension. Challenging attentional resources can influence how audiovisual speech information is perceived (Buchan & Munhall, 2012). Reducing extraneous cognitive load, such as minimizing distractions or simplifying the presentation of information, can lead to improved learning and performance (e.g., Bobis, Sweller & Cooper, 1993; Paas, Renkl, & Sweller, 2004). Based on these results, we posulate that individuals growing up in lower-noise situations may allocate more attention to integrating multimodal information, such as gestures, leading to a greater enhancement in degraded speech comprehension.

However, it is noteworthy that participants who reported higher levels of environmental noise demonstrated consistently strong performance in both clear and degraded speech conditions, indicating potential adaptive mechanisms to cope with auditory degradation. On the one hand, these findings suggest that reduced exposure to environmental noise may enhance the integration of visual cues, such as gestures, in the context of degraded speech comprehension. These results suggest that an individual's living environment, specifically the presence of environmental noise, can play a crucial role in shaping the effectiveness of multimodal communication strategies. On the other hand, these findings suggest that reduced exposure to environmental noise may impede the ability to compensate for degraded speech, necessitating the presence of additional visual cues.

Degraded speech comprehension is higher for individuals who experienced higher levels of environmental noise compared to individuals who experienced lower levels of environmental noise

Contrary to our hypotheses, individuals who were living in or had lived in high-noise environments found it easier to identify the verbs in the videos when speech was degraded than those who were living in or had lived in low-noise environments. These results suggest that participants accustomed to high levels of noise have developed robust cognitive mechanisms that enable them to effectively comprehend speech even under challenging conditions. While lower levels of environmental noise were advantageous for cognitive and linguistic abilities in earlier work, our findings suggest that individuals accustomed to noisy environments may possess heightened adaptation mechanisms that enable them to better comprehend speech when it is degraded. One explanation for this discrepancy is that previous work has solely studied unimodal language and reading abilities, whereas the current work studies more natural, multimodal language processing. Moreover, most behaviorally-relevant sounds (such as a person talking) occur against a background of everyday noise (e.g., traffic noise, music, other people talking etc.). Thus, adapting to noisy environments is a fundamental feature of the auditory system important for a range of listening conditions (Willmore et al., 2014; King and Walker, 2020). These findings thus contribute to the existing literature by shedding light on the complex interaction between environmental noise, multimodal communication, which often occurs in adverse listening conditions, and cognitive processing. Future research should further investigate the specific mechanisms underlying the relationship between these different factors, as it has implications for understanding individual differences in adapting to noisy environments and optimizing communication strategies.

Individual differences in audiovisual speech-gesture integration

Overall, these findings contribute to our understanding of how exposure to environmental noise affects how much we benefit from visual semantic information in noisy situations. By demonstrating that prior auditory experiences can affect audiovisual speech-in-noise comprehension and auditory processing, our results highlight the importance of considering the role of environmental noise in the study of audiovisual language comprehension. Our findings also have practical implications for the development of interventions to enhance communication in noisy environments. For example, interventions aimed at improving auditory processing and speech comprehension could incorporate training in the use of visual cues, such as gestures, to compensate for difficulties in processing auditory information. Furthermore, our results suggest that environmental noise mitigation strategies, such as implementing noise barriers and improving urban planning, could potentially have long-term benefits for cognitive functioning, including speech comprehension and audiovisual integration abilities.

Future research could extend the current findings by exploring the neural mechanisms underlying the observed effects of environmental noise on audiovisual language comprehension. Neuroimaging could be used to investigate how exposure to environmental noise influences neural activity involved in processing auditory and visual information during speech comprehension. Additionally, future studies could examine whether other factors, such as cognitive abilities, attention, or working memory capacity, affect the relationship between environmental noise exposure and the benefit of gestures during degraded speech comprehension.

Conclusion

We demonstrated that the amount of environmental noise in both current and previous living environments affected the degree to which individuals benefit from visual semantic information from gestures during degraded speech comprehension. Individuals who experienced lower levels of environmental noise in both their current and previous living environment benefitted more from gestures during degraded speech comprehension. Moreover, individuals who were living in or lived in high-noise environments found it easier to identify the verbs in the videos when speech was degraded than individuals who were living in or lived in low-noise environments. The differential effects observed between individuals with high and low levels of environmental noise exposure suggest that prolonged exposure to high-noise environments may lead to adaptive mechanisms that enhance speech processing in degraded conditions. This suggests a potential advantage for individuals living in noisy environments, at least in terms of their ability to decipher degraded speech, potentially due to enhanced perceptual and cognitive processes. Our findings highlight the importance of considering environmental noise exposure when investigating individual differences in audiovisual language comprehension. Future research should explore the underlying mechanisms driving these observed effects, such as attentional allocation, cognitive load, and neural adaptations, to provide a more comprehensive understanding of the complex interplay between environmental noise, multimodal cues, and speech comprehension in adverse listening conditions.

Data & code availability
The data and code used for the current study are available at https://osf.io/hn9us/

Funding
This work was supported by a Minerva Fast Track Fellowship from the Max Planck Society awarded to LD.

Ethics approval
The study was approved by the Faculty Ethics Committee of the Radboud University Nijmegen (approval code: ECSW-2020-049).

Acknowledgements
We would like to thank the Neurobiology of Language Department (Max Planck Institute for Psycholinguistics) for helpful discussions.

5 References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Bobis, J., Sweller, J., & Cooper, M. (1993). Cognitive load effects in a primary-school geometry task. *Learning and Instruction, 3*(1), 1–21. https://doi.org/10.1016/S0959-4752(09)80002-9

Buchan, J. N., & Munhall, K. G. (2012). The effect of a concurrent working memory task and temporal offsets on the integration of auditory and visual speech information. *Seeing and Perceiving, 25*(1), 87–106. https://doi.org/10.1163/187847611X620937

Bates, D., M. M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software.*

Drijvers, L., Jensen, O., & Spaak, E. (2021). Rapid invisible frequency tagging reveals nonlinear integration of auditory and visual information. *Human Brain Mapping, 42*(4), 1138–1152. https://doi.org/10.1002/hbm.25282

Drijvers, L., & Özyürek, A. (2017). Visual Context Enhanced: The Joint Contribution of Iconic Gestures and Visible Speech to Degraded Speech Comprehension. *Journal of Speech, Language, and Hearing Research: JSLHR, 60*(1), 212–222. https://doi.org/10.1044/2016_JSLHR-H-16-0101

Drijvers, L., Özyürek, A., & Jensen, O. (2018). Hearing and seeing meaning in noise: Alpha, beta, and gamma oscillations predict gestural enhancement of degraded speech comprehension. *Human Brain Mapping, 39*(5), 2075–2087. https://doi.org/10.1002/hbm.23987

Drijvers, L., van der Plas, M., Özyürek, A., & Jensen, O. (2019). Native and non-native listeners show similar yet distinct oscillatory dynamics when using gestures to access speech in noise. *NeuroImage, 194*, 55–67. https://doi.org/10.1016/j.neuroimage.2019.03.032

European Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise, 2002. Brussels, Belgium.

Evans, G. W., & Maxwell, L. (1997). Chronic Noise Exposure and Reading Deficits: The Mediating Effects of Language Acquisition. *Environment and Behavior, 29*(5), 638–656. https://doi.org/10.1177/0013916597295003

Hockey, G. R. J. (1970). Effect of loud noise on attentional selectivity. *The Quarterly Journal of Experimental Psychology, 22*(1), 28–36. https://doi.org/10.1080/14640747008401898

Holle, H., Obleser, J., Rueschemeyer, S.-A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage, 49*(1), 875–884. https://doi.org/10.1016/j.neuroimage.2009.08.058

Hygge, S., Evans, G. W., & Bullinger, M. (2002). A prospective study of some effects of aircraft noise on cognitive performance in schoolchildren. *Psychological Science, 13*(5), 469–474. https://doi.org/10.1111/1467-9280.00483

King, A. J., & Walker, K. M. (2020). Listening in complex acoustic scenes. *Current Opinion in Physiology, 18*, 63–72. https://doi.org/10.1016/j.cophys.2020.09.001

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lercher, P., Evans, G. W., & Meis, M. (2003). Ambient Noise and Cognitive Processes among Primary Schoolchildren. *Environment and Behavior, 35*(6), 725–735. https://doi.org/10.1177/0013916503256260

Maxwell, L. E., & Evans, G. W. (2000). THE EFFECTS OF NOISE ON PRE-SCHOOL CHILDREN'S PRE-READING SKILLS. *Journal of Environmental Psychology, 20*(1), 91–97. https://doi.org/10.1006/jevp.1999.0144

Moulin, A., Vergne, J., Gallego, S., & Micheyl, C. (2019). A New Speech, Spatial, and Qualities of Hearing Scale Short-Form: Factor, Cluster, and Comparative Analyses. *Ear and Hearing, 40*(4), 938–950. https://doi.org/10.1097/AUD.0000000000000675

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science, 32*(1/2), 1–8. http://www.jstor.org/stable/41953634

Rijksinstituut voor Volksgezondheid en Milieu. (2020). Atlas Leefomgeving - Verken en ontdek je leefomgeving. https://www.atlasleefomgeving.nl/kaarten?config=3ef897de-127f-471a-959b-93b7597de188&activateOnStart=layercollection&gm-x=155000.00000000003&gm-y=456478.71437812824&gm-z=2.894059846274022&gm-b=1544180834512%2Ctrue%2C1%3B1544715737496%2Ctrue%2C0.8

Seabi, J., Cockcroft, K., Goldschagg, P., & Greyling, M. (2012). The impact of aircraft noise exposure on South African children's reading comprehension: the moderating effect of home language. *Noise & Health, 14*(60), 244–252. https://doi.org/10.4103/1463-1741.102963

Thompson, R., Smith, R. B., Bou Karim, Y., Shen, C., Drummond, K., Teng, C., & Toledano, M. B. (2022). Noise pollution and human cognition: An updated systematic review and meta-analysis of recent evidence. *Environment International, 158*, 106905. https://doi.org/10.1016/j.envint.2021.106905

Willmore, B. D. B., Cooke, J. E., & King, A. J. (2014). Hearing in noisy environments: noise invariance and contrast gain control. *The Journal of Physiology, 592*(16), 3371–3381. https://doi.org/10.1113/jphysiol.2014.274886

Wilms, V., Drijvers, L., & Brouwer, S. (2022). The Effects of Iconic Gestures and Babble Language on Word Intelligibility in Sentence Context. *Journal of Speech, Language, and Hearing Research: JSLHR, 65*(5), 1822–1838. https://doi.org/10.1044/2022_JSLHR-21-00387