RESEARCH NOTES

Author: Gerard Llorach
Webpage: https://gerardllorach.weebly.com/

ENRICH PROJECT

http://www.enrich-etn.eu/
Funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 675324 (ENRICH).

Disclaimer

The information contained in this document is for general information purposes only. The author assumes no responsibility for errors or omissions in the contents. In no event shall the author be liable for any special, direct, indirect, consequential, or incidental damages or any damages whatsoever, whether in an action of contract, negligence or other tort, arising out of or in connection with the use of the contents of this document. The author reserves the right to make additions, deletions, or modification to the contents at any time without prior notice.

External links disclaimer. This document may contain links to external websites that are not provided or maintained by or in any way affiliated with the author. Please note that the author does not guarantee the accuracy, relevance, timeliness, or completeness of any information on these external websites.

IMPACT OF AUDIOVISUAL INTEGRATION ON ACOUSTIC COMM. IN HI ADULTS

Visual cues are a highly important component in communication (Campbell 2008, Holler *et al.* 2009; Lee and Noppeney 2011; Wagner *et al.*, 2014; Drijvers and Özyürek, 2016; Gil-Carvajal *et al.*, 2016). Examples of how visual cues interact with speech information include:

- Phonological information, which is conveyed by the visual speech cues (lips, tongue and teeth movements, mouth opening).
- Prosodic information, which is conveyed by pauses, head inclinations, gaze, rhythmic hand and body movements, and eyebrow movements.
- Semantic content, which is conveyed by iconic gestures, and visual text.
- Environmental information, which provides room information and externalization cues.

However, the degree of benefit differs greatly across listeners, presumably in relation to variables including multisensory integration ability, age, cognitive ability, severity and duration of hearing loss, acoustic environments, integrity of visual cues, and the spatiotemporal coherence of auditory and visual cues (*e.g.,* Summerfield, 1992; Rajj et al., 2000; Erber, 2003; Alsius et al., 2005; Navarra et al., 2005; Spence, 2007; Feld and Sommers 2009; Obermeier *et al.* 2012; Altieri and Hudock, 2014; Rudner *et al.* 2015; Krull and Humes, 2016).

Unanswered questions:
· How is the ability to integrate audiovisual information affected by distortion of visual cues and/or distortion of the acoustic signal introduced by a hearing aid?
· What are the effects of hearing loss and hearing-aid processing on the ability to integrate audiovisual phonological, prosodic, semantic, or environmental information?
· Do unaided and aided hearing-impaired listeners adopt specific strategies to benefit from audiovisual integration?
· How does age-related hearing and vision decline affect this strategy?
· Does cross-modal reorganization influence audiovisual integration abilities in hearing-impaired listeners (hearing impairment causes reorganization in the auditory and visual cortices)?

WHAT REMAINS TO BE DONE WITH HMD

EEG + HMD (Cattan et al. 2018) and visual cues → compare video, VC and AO with EEG signals
HMD + Virtual FOV → increasing virtual FOV gets ecological head behavior?

DIFFERENCES BETWEEN 2D AND HMD (evaluation)

Maxime Cordeil, Tim Dwyer, Karsten Klein, Bireswar Laha, Kim Marriott, Bruce Thomas. **2016**. Immersive Collaborative Analysis of Network Connectivity: CAVE-style or Head-Mounted Display? IEEE Transactions on Visualization and Computer Graphics
10.1109/TVCG.2016.2599107
http://ieeexplore.ieee.org/document/7539620/
Collaborative analysis. Few questions asked. Some open feedback given at the end.

C. J. Fluke, D. G. Barnes. **2016**. The Ultimate Display
https://arxiv.org/abs/1601.03459
Astronomical datasets

Adam Philpot, Maxine Glancy, Peter J. Passmore, Andrew Wood, and Bob Fields. **2017**.User Experience of Panoramic Video in CAVE-like and Head Mounted Display Viewing Conditions. In Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video (TVX '17).
https://dl.acm.org/citation.cfm?id=3077550
360º videos. Bee man scene plus two others.
In a study on anxiety [14] showed that anxiety produced by fear of heights was higher in a CAVE compared to in a HMD
In [15] performance on a modified Stroop task under low and high stress conditions, was compared for desktop, HMD and CAVE-like viewing conditions. The CAVE-like system induced the greatest sense of presence.
high or low emotional impact videos, viewed on a HMD compared to a tablet, affected the viewer's pro-environmental attitude and behaviour [5]. They found that the increased immersion and higher emotional impact of the HMD appeared to increase self-reported measures, but that higher immersion did not increase pro-environmental behaviour significantly
None of the users exhibited any signs of cyber sickness, though two users mentioned that they felt slightly disoriented when removing the HMD.
Comfort not valid, as in the CAVE they were sitting down.
Some users found the HMD quite uncomfortable due to its weight pressing on the face, some eye strain due to the closeness of the screen to the eye and feeling slightly claustrophobic.
Attention equal. Fear of missing out if looking somewhere else in HMD and CAVE.

Doug A. Bowman , Ameya Datey , Young Sam Ryu , Umer Farooq , Omar Vasnaik. Empirical Comparison of Human Behavior and Performance with Different Display Devices for Virtual Environments (**2002**)
http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.103.2623
Turns (natural, manual, combined) with HMD and three screen display (CAVE, SID) -> watch out, it is from 2002!!! Only a 50% preference!

Sarah Sharples, Sue Cobb, Amanda Moody, and John R. Wilson. **2008**. Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. Displays 29, 2 (mar 2008), 58–69. DOI: http://dx.doi.org/10.1016/j.displa.2007.09.005
http://www.sciencedirect.com/science/article/pii/S014193820700100X?via%3Dihub

Four display conditions: HMD, desktop, projection screen and curved large screen (reality theater). SSQ scores higher for HMD vs 2D and vs RT.

CAVE versus Head-Mounted Displays: Ongoing thoughts. **2017**
http://www.ingentaconnect.com/contentone/ist/ei/2017/00002017/00000003/art00006
Not interesting, just surface.

User Experience and Immersion of Interactive Omnidirectional Videos in CAVE Systems and Head-Mounted Displays. **2017**
https://link.springer.com/chapter/10.1007/978-3-319-68059-0_20
Users did not try both, so its not interesting. Speaks a lot about immersion and presence, which cannot be compared if they didn't try both.

Rand, D., Kizony, R., Feintuch, U., Katz, N., Josman, N., Rizzo, A.S. and Weiss, P.L., **2005**. Comparison of two VR platforms for rehabilitation: video capture versus HMD. Presence: Teleoperators & Virtual Environments, 14(2), pp.147-160.
https://www.gesturetekhealth.com/sites/default/files/general/4.2.7_rand_presence_proof_0.pdf
"In contrast, in the Virtual Office, a significant main effect for age was found; the **older group preferred the HMD** as compared to the younger group. This finding was unexpected since some of the elderly participants reported difficulty in focusing on the images while wearing the HMD, a problem that was particularly noticeable for the participants who wore bifocal glasses."
"The decreased performance of the older group in both types of platforms points to the validity of the tasks performed within these VR systems since we expect to see differences as a function of age. A decline in visual acuity (e.g., Stanney et al., 1998) as well as perceptual-cognitive abilities such as memory (e.g., Josman & Hartman-Maeir, 2000) could lead to the decreased performance demonstrated in this study. The performance of the older group may also be influenced by the task characteristics"

STATEMENTS

Philpot 17: Also, in terms of physiology, in the CAVE, vision is normal, whereas for the HMD, **accommodation and vergence are decoupled, leading to possible discomfort**: David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue

Philpot 17: **a significant preference for real turns in the HMD, and for virtual turns in the SID (Spatially Immersive Display) The user wears stereo glasses and a head tracker****: Doug A Bowman, Ameya Datey, Young Sam Ryu, Umer Farooq, and Omar Vasnaik. 2002. Empirical Comparison of Human Behavior and Performance with Different Display Devices for Virtual Environments. * from 2002 ** missing back wall and top and bottom.

Philpot 17: (impressions) **embodiment** Some users mentioned that they felt disembodied, more so when using the HMD. The fact that when they looked down and did not see their legs or any kind of body representation was quite strange for some users. **Size of image**: bigger in CAVE. **peripheral vision**: liking of image everywhere in CAVE, Nevertheless distracting, also in HMD. **Cube effects of CAVE**.

Philpot 17: **comfort**: more in CAVE, as they didn't have to use a headset (they were sitting on a sofa in CAVE and chair in HMD).

Philpot 17: **higher presence in HMD**. Contradictory with [15] in Philpot. **Immersion and presence are decisive on preference**.

MEASUREMENTS OF HEAD MOTION

Willard R. Thurlow, John W. Mangels and Philip S. Runge. **1967**. Head Movements During Sound Localization. J. Acoust. Soc. Am. 42, 489
http://asa.scitation.org/doi/abs/10.1121/1.1910605
Classification of the move according to axis (type 1 = only yaw, type 2 = yaw and pitch, etc…)
Maximal head rotation angle threshold: >10º
Minimal head rotation angle: >3º
They did not mention how they classified the movements in respect of time, but maximal to 5 - 6 types of movement in 5 seconds.
· **Maximal head rotation** (in our case over the whole story-condition)

Brimijoin, W. O., McShefferty, D., & Akeroyd, M. A. (**2012**). Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task. Hearing Research, 283(1-2), 162–168. http://doi.org/10.1016/j.heares.2011.10.009
2-3 seconds of sentences. Target always in front, distractor on sides.
Plots histogram of bins and median of either yaw, pitch or roll. Two way ANOVA on the median?
· **Yaw, pitch** (in our case, for each target speaker position?)
They plot maximal SNR, but they don't do any statistical tests, just comment it on the discussion.
Pitch: "All listeners consistently pitched their heads downwards during the task"

Grange, J. A. , and Culling, J. F. (**2016**). “ The benefit of head orientation to speech intelligibility in noise,” J. Acoustic. Soc. Am. 139, 703–712.
https://doi.org/10.1121/1.4941655
Average unsigned head orientation corresponds to amount of movement. Listeners' head orientation strategies turned out to be too heterogeneous for us to be able to distinguish between the potential strategies listed in the introduction.
More than 10º away from speech is considered to turn away from speech (56% of subjects in their study).

Grange, J.A., Culling, J.F., Bardsley, B., Mackinney, L.I., Hughes, S.E. and Backhouse, S.S., **2018**. Turn an Ear to Hear: How Hearing-Impaired Listeners Can Exploit Head Orientation to Enhance Their Speech Intelligibility in Noisy Social Settings. Trends in hearing, 22, p.2331216518802701.
https://journals.sagepub.com/doi/pdf/10.1177/2331216518802701
Visual stimuli at 0º. 64% of the YNH subjects turned the head away. Audiovisual experiment from Grange.

Gifford, René H., Louise Loiselle, Sarah Natale, Sterling W. Sheffield, Linsey W. Sunderhaus, Mary S. Dietrich, and Michael F. Dorman. "Speech understanding in noise for adults with cochlear implants: effects of hearing configuration, source location certainty, and head movement." Journal of Speech, Language, and Hearing Research 61, no. 5 (**2018**): 1306-1321.
https://pubs.asha.org/doi/pdf/10.1044/2018_JSLHR-H-16-0444?casa_token=6cWKqfl5pRAAAAAA:pBw8P0eWS9Aje4CScR3eZ_clih57poqBOnx8676EdwuqvkqVHmbtzCJM6w_JJZlO0cy7r6Xiojeak_Yp
No audiovisual. No measures of head turn, only allowed and not allowed. "all listeners commented that the task was much simpler when allowed to use head turns.". "investigated the effects of head turn behaviors on speech understanding with roving sources"
"We found no effect of source location certainty nor head turns on sentence recognition in noise for any of the source azimuths or listening configurations. The former was consistent with Davis et al. (2016) but in contrast

to two previous studies demonstrating a significant benefit for source location certainty in listeners with normal hearing (Brungart & Simpson, 2007; Kidd et al., 2005)."

Yi Shen, Monica L. Folkerts, and Virgina M. Richards. **2017**. Head movements while recognizing speech arriving from behind
http://asa.scitation.org/doi/suppl/10.1121/1.4976111
CRM (two speakers or one speakers and noise). Two sentences for trial consecutively played from the back. (165º and -165º). No improvement of SNR with the head movement.
Average of 400ms of head orientation at the end of each trial.
>10º movement classifies HT and NHT (average over trial)
Negative or positive direction of movement towards target. A head-turn is a movement from the initial orientation, but doesn't specify the axis.
Head turn latency of 1s after target signal

Lu, H., McKinney, M., Zhang, T. and Oxenham, A.J., **2018**. Tracking eye and head movements in natural conversational settings: Effects of hearing loss and background noise level. The Journal of the Acoustical Society of America, 143(3), pp.1743-1743.
"Preliminary results: younger listeners tended to use their eyes, as opposed to head movements, more than the older listeners."
No differences between hearing type. "younger listeners, older listeners with clinically normal hearing, and older listeners with mild-to-moderate hearing loss."

HEAD MOVEMENT AND LISTENING

Hadley, L.V., Whitmer, W.M., Brimijoin, W.O. and Naylor, G., 2021. Conversation in small groups: Speaking and listening strategies depend on the complexities of the environment and group. Psychonomic Bulletin & Review, 28(2), pp.632-640.
https://link.springer.com/article/10.3758/s13423-020-01821-9
Conversations dyadic and triadic. Head angle was less directly oriented to the speaker in babble than speech-shaped noise, and was not affected by noise level.
When analysing listeners' head orientations (Figs. 1d and 1e) there was **no effect of noise level** ($F(2.91, 90.15) = 2.24$, $p = .090$).

Hadley, L.V., Brimijoin, W.O. and Whitmer, W.M., **2019**. Speech, movement, and gaze behaviours during dyadic conversation in noise. Scientific reports, 9(1), p.10451.
https://www.nature.com/articles/s41598-019-46416-0
More ecologically valid. Needs further reading. Main findings: leaning forward as noise increases; looking more at the mouth as noise increases. No head-orientation SNR benefit.

Simpson, W.E. and Stanton, L.D., **1973**. Head movement does not facilitate perception of the distance of a source of sound. The American journal of psychology, pp.151-159.
Azimuth localization → head movements helps
Elevation localization → head movements helps or not (Wallach 1940 vs Roffler and Butler 1968)

Shen, Y., Folkerts, M.L. and Richards, V.M., **2017**. Head movements while recognizing speech arriving from behind. The Journal of the Acoustical Society of America, 141(2), pp.EL108-EL114.
Same as Grange? "These observed head movements are not consistent with a strategy that maximizes either the target level at a single ear or binaural unmasking of speech." 10 YNH subjects.

Woestmann, M., Schmitt, L.M. and Obleser, J., **2019**. Does closing the eyes enhance auditory attention? Eye closure increases attentional alpha-power modulation but not listening performance. Journal of cognitive neuroscience, pp.1-14.
Eyes closed → more concentration. Looking down is not spoken about.

Van Barneveld, D.C., Van Grootel, T.J., Alberts, B. and Van Opstal, A.J., **2011**. The effect of head roll on perceived auditory zenith. Experimental brain research, 213(2-3), pp.235-243.
https://link.springer.com/content/pdf/10.1007%2Fs00221-011-2741-9.pdf
The only paper I found that talks about roll. "Subjects sat either upright, or with their head left/right rolled sideways by about 35 relative to gravity, whilst judging whether a broadband sound was heard left or right from the head-centred or world-centred zenith."

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (**2010**). Variability of eye movements when viewing dynamic natural scenes. Journal of vision, 10(10), 28-28.
https://jov.arvojournals.org/article.aspx?articleid=2121333
"We conclude that the stimuli types often used in laboratory experiments, static images and professionally cut material, are not very representative of natural viewing behavior." By natural viewing behavior they refer to watching videos of outdoor scenes.

MEASUREMENTS OF HEAD MOTION (SYNTHESIS)

Le, B.H.; Xiaohan Ma; Zhigang Deng, **2012** Live Speech Driven Head-and-Eye Motion Generators, in *Visualization and Computer Graphics, IEEE Transactions on*
http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6165277&navigation=1
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6165277
https://www.youtube.com/watch?v=VZF4mSilYGM
(Dyadic conversation scenario) Three euler angles, velocity magnitude and acceleration magnitude.

Sergey Levine, Philipp Krähenbühl, Sebastian Thrun, and Vladlen Koltun. **2010** Gesture controllers. In ACM SIGGRAPH 2010 papers.
http://dl.acm.org/citation.cfm?doid=1833349.1778861
(Body parts, not head precisely) Temporal extent, spatial extent, velocity, acceleration, curvature, height

Yilong Liu, Feng Xu, Jinxiang Chai, Xin Tong, Lijuan Wang, and Qiang Huo. 2015. Video-audio driven real-time facial animation. ACM Trans. Graph. 34, 6, Article 182 (October 2015), 10 pages. DOI:
https://doi.org/10.1145/2816795.2818122
Head pose (I guess orientation)

GAZE - EOG FOR HEARING AIDS

Ľuboš Hládek, Bernd Porr, W. Owen Brimijoin. **2017**. Real-time estimation of horizontal gaze angle by saccade integration using in-ear electrooculography. PLoS ONE 13(1)
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0190420
Some authors have suggested that using eye gaze angle to steer hearing aid directional microphones could be of benefit to a listener [4,5] (Kidd 2013 and Hart 2009).

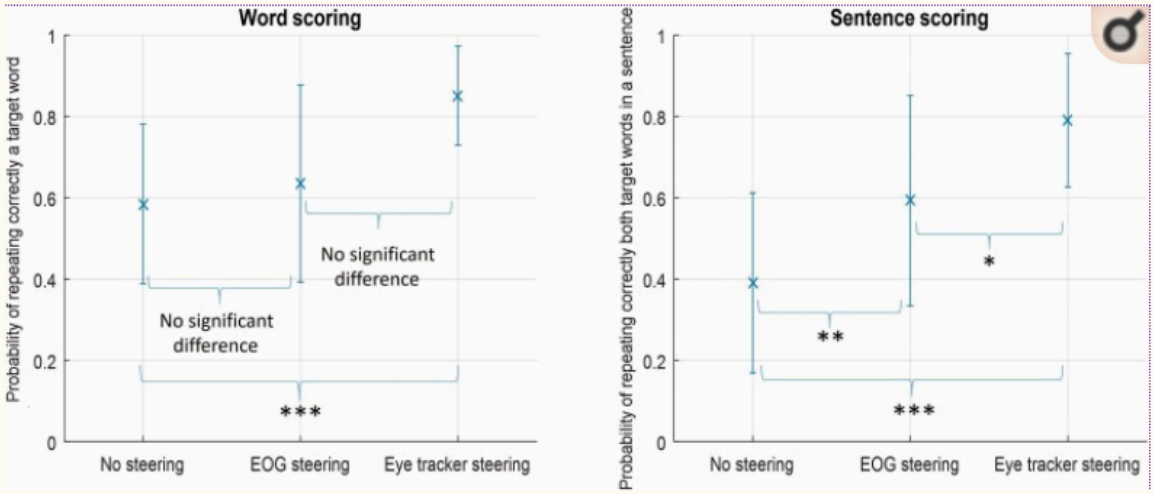
2019. On the Interaction of Head and GazeControl With Acoustic Beam Width of a Simulated Beamformer in a Two-Talker Scenario
https://www.researchgate.net/publication/335993816_On_the_Interaction_of_Head_and_Gaze_Control_With_Acoustic_Beam_Width_of_a_Simulated_Beamformer_in_a_Two-Talker_Scenario
From the Hládek, lab experiments.

Kidd G Jr., Favrot S, Desloge JG, Streeter TM, Mason CR. Design and preliminary testing of a visually guided hearing aid. J Acoust Soc Am. Boston, MA, United States.: Acoustical Society of America; **2013**; 133: EL202–EL207. https://doi.org/10.1121/1.4791710 PMID: 23464129
Not tested with users, only a KEMAR

Hart J, Onceanu D, Sohn C, Wightman D, Vertegaal R. The Attentive Hearing Aid: Eye Selection of Auditory Sources for Hearing Impaired Users. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2009. pp. 19–35. 10.1007/978-3-642-03655-2_4
https://link.springer.com/content/pdf/10.1007/978-3-642-03655-2_4.pdf
52" plasma screen. 14 hearing-impaired individuals. Selecting sources by pointing, buttons or eye tracker. Three male actors each reading the nine passages, for a total of 27 movie files.

Cattan, G., Andreev, A., Mendoza, C. and Congedo, M., **2018**, April. The Impact of Passive Head-Mounted Virtual Reality Devices on the Quality of EEG Signals. In Workshop on Virtual Reality Interaction and Physical Simulation.
https://hal.archives-ouvertes.fr/hal-01774915
Andreev A. contacted me on linkedin. Interesting for EEG+HMD+hearing aid. He is also doing research with ECAs (https://alucard321.itch.io/companion)

Favre-Felix, A., Graversen, C., Hietkamp, R.K., Dau, T. and Lunner, T., **2018**. Improving speech intelligibility by hearing aid eye-gaze steering: Conditions with head fixated in a multitalker environment. Trends in hearing, 22, p.2331216518814388.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6291882/
Other literature mentioned in the paper: "The idea of using eye gaze to steer a hearing aid has already been explored in several previous studies (Hart, Onceanu, Sohn, Wightman, & Vertegaal, 2009; Kidd, 2017; Kidd, Favrot, Desloge, Streeter, & Mason, 2013)."


Word scoring
Probability of repeating correctly a target word
1
0.8
0.6
0.4
0.2
0
No significant difference
No significant difference
***
No steering
EOG steering
Eye tracker steering
Sentence scoring
Probability of repeating correctly both target words in a sentence
**
*
***
No steering
EOG steering
Eye tracker steering

LOUDNESS PERCEPTION

Smeds, Karolina, Gitte Keidser, Justin Zakis, Harvey Dillon, Arne Leijon, Frances Grant, Elizabeth Convery, and Christopher Brew. "Preferred overall loudness. II: Listening through hearing aids in field and laboratory tests" International Journal of Audiology 45, no. 1 (**2006b**): 12-25.
https://www.tandfonline.com/doi/abs/10.1080/14992020500190177?casa_token=fLIPpcQNjnwAAAAA:N41WglDWLCotuB1F5cQtIGfFcOUI2Kx2DcL3w9Ju8tVYSHZWxzbTVw1mafUOy6eWXepnJ4mqnnMWWw
https://www.tandfonline.com/doi/pdf/10.1080/14992020500190177?needAccess=true
"In subsequent laboratory tests, using the same hearing aid, both hearing-impaired and normal-hearing listeners preferred less than normal overall calculated loudness, and larger reductions for higher input levels."
Field versus Lab comparison.
Hearing impaired prefer lower loudness, so the hearing aid should not restore completely the loudness.
Laboratory is video with two loudspeakers behind a curtain.

Smeds 2006a "The loudspeakers were hidden behind curtains that framed the television screen so that the participants would not focus visually on anything other than the screen. There was a special point in hiding the loudspeakers, because there is research which suggests that loudness perception is biased by the visual distance to the loudspeakers (Mershon et al, 1981)."

No loudness binaural summation: "[...] loudness summation could have been adjusted. However, this was not necessary for any of the participants."

Mershon, D.H., Desaulniers, D.H., Kiefer, S.A., Amerson, T.L.J. & Mills, J.T. **1981**. Perceived loudness and visually-determined auditory distance. Perception, 10, 531/43
Smeds 2006a "there is research which suggests that loudness perception is biased by the visual distance to the loudspeakers (Mershon et al, 1981)."

Keidser, G., Katsch, R., Dillon, H. and Grant, F., **2000**. Relative loudness perception of low and high frequency sounds in the open and occluded ear. The Journal of the Acoustical Society of America, 107(6), pp.3351-3357.
Smeds 2006b: "There is research indicating that loudness perception might be influenced by the way sound is presented to the listener (Keidser et al, 2000)."
"measurements are done in an occluded ear than when they are done in an open ear, even though all measurements are expressed as dB SPL at the eardrum" → Similar to Florian Denk medi seminar (gap between loudspeaker and headphone presentation)
" The results show that an average of 10 dB higher level was selected for the 500-Hz octave band when listening with the occluded ear than when listening with the open ear"

Oetting, D., Bach, J.H., Krueger, M., Vormann, M., Schulte, M. and Meis, M., **2019**. Subjective loudness ratings of vehicle noise with the hearing aid fitting methods NAL-NL2 and trueLOUDNESS. In Proceedings of the International Symposium on Auditory and Audiological Research (Vol. 7, pp. 289-296).
https://proceedings.isaar.eu/index.php/isaarproc/article/view/2019-33
Experiment with hearing impaired for the field experiment.
"The EuroTrak survey in Germany(EuroTrak DE, 2018)showed that the dimension “comfort with loud sounds”was the most important criterion for the overall satisfaction with hearing aids"

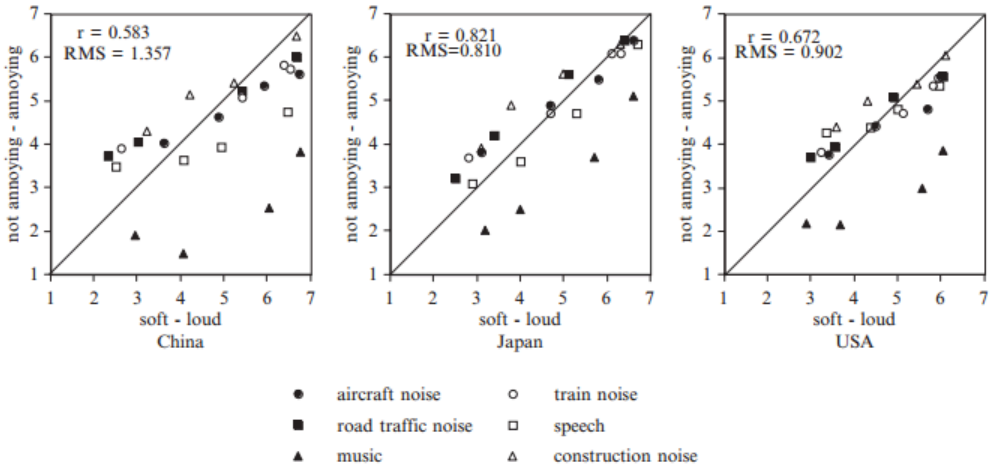
Hugo Fastl, **2004**, Audio-visual interactions in loudness evaluation. ICA 2004 Kyoto, Intern. Commision on Acoustics
http://mediatum.ub.tum.de/doc/1138383/152725.pdf
Red/Green train, how the levels were compared? By recall? Or adaptive methods?
It is a summary of the work by Christine Patsouras
https://www.amazon.de/-/en/dp/3832219285/ref=sr_1_1?dchild=1&qid=1586166549&refinements=p_27%3AChristine+Patsouras&s=books&sr=1-1&text=Christine+Patsouras

Hugo Fastl, Mary Florentine. **2010**. Loudness in daily environments. Volume 37 of the series Springer Handbook of Auditory Research pp 199-221
https://link.springer.com/chapter/10.1007/978-1-4419-6712-1_8
We find that some assumptions made in the laboratories do not apply to daily environments.
When visual stimuli are presented together with acoustic stimuli in experiments that are designed to start to bridge the gap between laboratory experiments and our perceptions in daily environments, audio–visual interactions can be observed. **In general, these effects are small, but significant**.

"Audio–visual interactions have also been observed under more ecologically valid conditions (e.g., Fujiwara et al. 2006)"

There is no perfect method for measuring loudness. Loudness researchers need to choose the best measurement method from what is available, while keeping in mind its limitations.

Acceptable methods for measuring loudness provide data conforming to two principles. The first is an ordinal indicant of relative loudness. The second principle is that loudness equalities must be transitive: If acoustic signal A1 is as loud as signal A2, and A2 is as loud as A3, then A1 must be as loud as A3.
equal-loudness measurements continue to be the "gold" standard to which results obtained by other methods must conform.



However, the decrease in perceived loudness is about the same magnitude for still vs. moving pictures (Fastl 2004). An even more realistic situation for the listener exists when sounds and moving pictures recorded at the passenger seat of a driving car are reproduced in a car simulator (Fastl 2004). In these situations, the perceived loudness is decreased by about 15% compared to judgments of the sound without the visual stimuli. However, for some individual listeners, the audio–visual interaction can lead to a loudness reduction of even more than 50%, whereas for others the effect is very small and only a few percent.

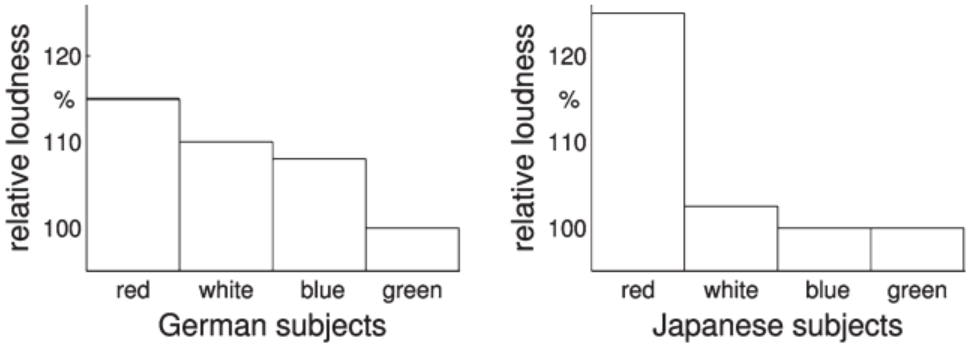


Fig. 8.3 Relative loudness ratings for trains of different colors by German subjects (*left*) and Japanese subjects (*right*)

Interesting: In other words, the barriers that reduced the visibility of the sound source – without entirely eliminating it – reduced the apparent loudness. However, the barrier that totally obscured the sight of the sound source increased the apparent loudness. expectation can influence perception.

Parizet, E. and Koehl, V., **2011**. Influence of train colour on loudness judgments. Acta Acustica united with Acustica, 97(2), pp.347-349.
https://hal.univ-brest.fr/file/index/docid/585433/filename/paper_acta_trains.pdf
"Patsouras et al.[5] or Rader et al.[8] in which subjects had to evaluate the loudness of a passing-by train noise while watching pictures of the train, the colour of which being varied. In contrast to these studies, **no influence of train colour on loudness evaluation** could be found."
"This discrepancy could be due to cultural effects (French subjects vs. German or Japanese ones). Such a cultural effect has already been found in some other topics (e.g. booming noise of a car [11], or the negative or neutral connotation of the word "loudness" [12]). But such differences were mainly found between Japanese and European subjects – the perception within Europe being homogeneous."

Arieh, Y. and Marks, L.E., **2011**. Measurement of loudness, Part II: Context effects. In Loudness (pp. 57-87). Springer, New York, NY.
https://link.springer.com/chapter/10.1007/978-1-4419-6712-1_3
"review and summarize current understanding of the ways that context affects loudness".
"discrete suprathreshold sounds, usually of relatively short duration"
Mostly theoretical, not related to real-life. From the same book ("Loudness") as Fastl 2010.

Fujiwara, M., Aono, S. and Kuwano, S., **2006**, December. Audio-visual interaction in the image evaluation of the environment-An on-site investigation. In INTER-NOISE and NOISE-CON Congress and Conference Proceedings (Vol. 2006, No. 6, pp. 1213-1220). Institute of Noise Control Engineering.
https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Audio-visual+interaction+in+the+image+evaluation+of+the+environment-An+on-site+investigation&btnG=
Paper not found!
Fastl 2010: "Audio–visual interactions have also been observed under **more ecologically valid conditions** (e.g., Fujiwara et al. 2006). Fujiwara et al. (2006) took participants in an experiment to a newly developed residential town in Japan where they were asked to judge the auditory and visual environment. Their judgments were made using a semantic differential method (Osgood et al. 1957, see also Chap. 1) in which

participants rated their impressions on adjective scales to obtain information on the meaning of the perception. They were tested in three conditions: auditory environment alone, visual environment alone, and overall (auditory and visual) environment. Results indicate that for the adjective pair “quiet–noisy” **the addition of a visual environment to the auditory environment results in judgments of a quieter overall environment** than the auditory environment alone."

Florentine, M., **2014**. A Quarter-Century’s Perspective on a Psychoacoustical Approach to Loudness. In Perspectives on Auditory Research (pp. 125-147). Springer, New York, NY.
https://link.springer.com/content/pdf/10.1007%2F978-1-4614-9102-6_8.pdf
Binaural loudness summation appears for tones but not for audiovisual speech. Ecologically valid setups.
"binaural-to-monaural loudness ratio"
"These results indicate that the loudness of a visually present talker in daily environments is little affected by switching between binaural and monaural listening."
"Michael Epstein and I have formally studied this phenomenon, which is known as Binaural Loudness Constancy, BLC—an almost complete absence of binaural loudness summation."

Epstein, M. and Florentine, M., **2012**. Binaural loudness summation for speech presented via earphones and loudspeaker with and without visual cues. The Journal of the Acoustical Society of America, 131(5), pp.3981-3988.
https://asa.scitation.org/doi/abs/10.1121/1.3701984?casa_token=i4aiaUdGm7oAAAAA:1UypFgDGsRdmz6RaRk0yiSAQ20B25zN5JN5F9ihO-AkVD2lHX1j8tVAS3FTfEw_6DqLyz7foUcyvSQ
More precise description of Florentine 2014. She found that in the classrooms, the kids did not find a difference in loudness when occluding an ear. "This supports the phenomenon of binaural loudness constancy and underscores the importance of ecological validity in loudness research"

Jeon, J.Y. and Jo, H.I., **2018**. Three-dimensional virtual reality-based subjective evaluation of road traffic noise heard in urban high-rise residential buildings. Building and Environment.
https://www.sciencedirect.com/science/article/pii/S0360132318306929
Probably a good introduction to the experiment I will do. Loudness annoyance, but results are not presented very clearly. Using HMD and some references to other studies with HMDs, but mostly not related to loudness. Loudness shown in a room. Combinations of HRTFs and HMD. "The results of this study confirm the applicability and necessity of an HRTF and HMD when evaluating noise in residential buildings."

Preis, A., Hafke-Dyx, H., Szychowska, M., Kocinski, J. and Felcyn, J., **2016**. Audio-visual interaction of environmental noise. Noise Control Engineering Journal, 64(1), pp.34-43.
https://www.researchgate.net/profile/Jan_Felcyn/publication/295860313_Audio-visual_interaction_of_environmental_noise/links/56d4487e08ae2ea08cf8e904/Audio-visual-interaction-of-environmental-noise.pdf
A better introduction to noise in audiovisual. Only still images and mainly mismatch of noise source and environment shown. Noise annoyance formulas.
"Maffei et al.2 analyzed how the visual characteristics of a screen can influence the noise perception of local residents. In the case of transparent barriers where sound sources were seen, perceived annoyance, as well as perceived loudness, were judged lower than when compared to the situations in which the sound sources were not seen (industrial, opaque or green barriers, see also Aylor and Marks3)"
"a nice image of a car drastically reduced the negative loudness assessment (Hashimoto and Hatano13) or unpleasantness (Hatano et al.14)."

Aylor, D.E. and Marks, L.E., **1976**. Perception of noise transmitted through barriers. The Journal of the Acoustical Society of America, 59(2), pp.397-400.
Preis et al. 2016: "In the case of transparent barriers where sound sources were seen, perceived annoyance, as well as perceived loudness, were judged lower than when compared to the situations in which the sound sources were not seen (industrial, opaque or green barriers, see also Aylor and Marks)"
Arieh and Marks 2011: "the more solid the barrier, and hence the less visible the sound source, the louder the sounds are judged (Marks and Aylor 1976)."

Zahorik, P. and Wightman, F.L., **2001**. Loudness constancy with varying sound source distance. Nature neuroscience, 4(1), p.78.
At-ear loudness is different from distanced loudness with same loudness levels? Concept learned at ICA 2019. Florentine 2014 and Epstein talk about it.

Keidser, G., O'Brien, A., Carter, L., McLelland, M. and Yeend, I., **2008**. Variation in preferred gain with experience for hearing-aid users. International Journal of Audiology, 47(10), pp.621-635.
https://www.tandfonline.com/doi/full/10.1080/14992020802178722?casa_token=_8iRENxBjtUAAAAA%3AkMbY1c0yU0ccmTO25vZwP2sqMiMVCVSsk7lsYdMgVRkRw1EjfbabZctEf1gtbMQsT3OrY1ULjUI
Audiometry does not account for loudness and needs to be managed over time.

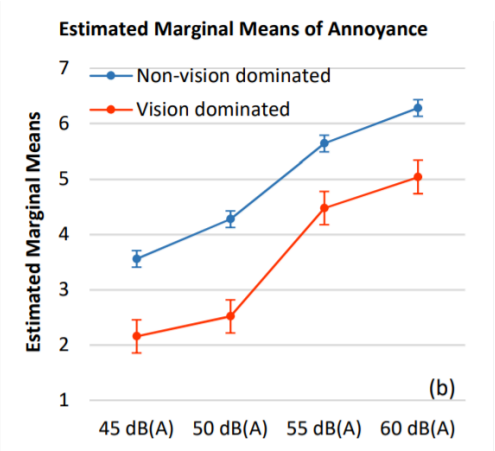
Beach, E.F., Williams, W. and Gilliver, M., **2012**. The objective-subjective assessment of noise: Young adults can estimate loudness of events and lifestyle noise. International Journal of Audiology, 51(6), pp.444-449.
https://www.tandfonline.com/doi/full/10.3109/14992027.2012.658971?casa_token=mSobO9TPE50AAAAA%3ASpO83DMCM-obSg4VKF9KIDcjPoFYBQNeW8HjQ6DCEhQtFtblh7Bsay2fKw-RsccPtesMbEnaTjNgKA
Wearing a loudness measurement device and writing down perceived loudness. "individuals can make a reasonable estimate of the loudness of events they experience and the overall level of noise they experience"

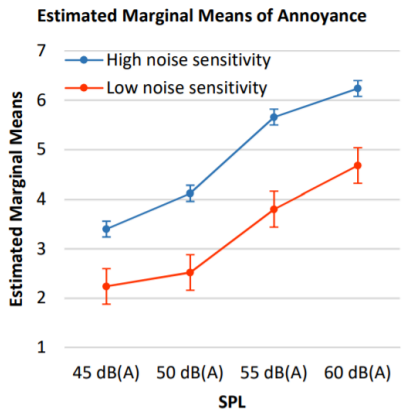
CONCLUSIONS

· Reduced visibility of the sound source - less loudness. Totally obscured the sight of the sound source - more loudness.
· Visible sound sources → lower annoyance (Sun et al. 2018) and lower loudness (Aylor and Marks 1976).
· The decrease in perceived loudness is about the same magnitude for still vs. moving pictures (Fastl 2004).

ANNOYANCE

Sun, K., De Coensel, B., Sanchez, G.M.E., Van Renterghem, T. and Botteldooren, D., **2018**. Effect of interaction between attention focusing capability and visual factors on road traffic noise annoyance. Applied Acoustics, 134, pp.16-24.
https://biblio.ugent.be/publication/8578820/file/8578823.pdf
75 participants. Mock-up room with a screen (as a window). Different videos, only one with audiovisual match (in the others the sound sources cannot be seen). Visibility of sources is an important factor, it reduces annoyance. Individual noise sensitivity and vision dominated. Interesting section (2.5 Audiovisual aptitude and noise sensitivity assessment).
"t the effect of being a vision dominated listener is almost as significant as the effect of noise sensitivity – a known stable personality trait – but more importantly, this personal factor interacts with visual factors. This factor should therefore be considered in future investigations."
"it has been shown that other features such as frequency and temporal content, sharpness and loudness also change people's preference towards sounds."

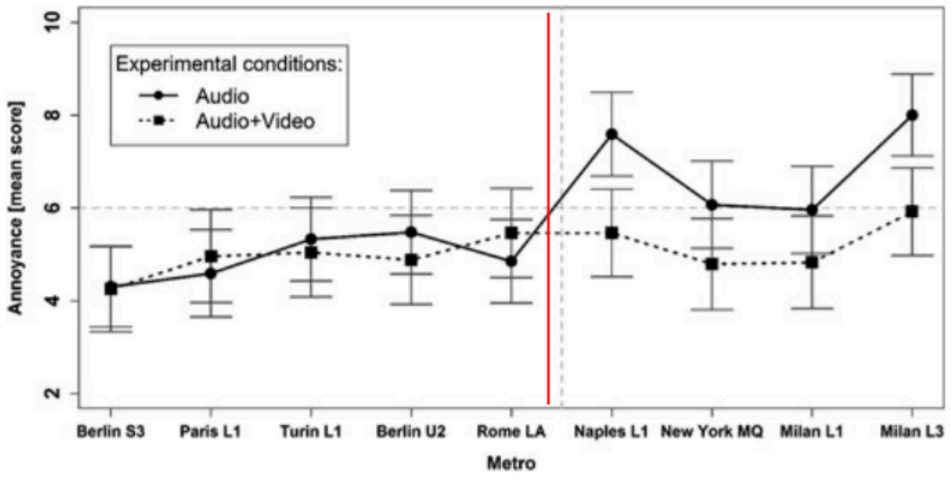




Schreckenberg, D., Belke, C., & Spilski, J. (**2018**). The Development of a Multiple-Item Annoyance Scale (MIAS) for Transportation Noise Annoyance. International journal of environmental research and public health, 15(5), 971.
Good introduction to annoyance definitions and scales.

Iachini, T., Maffei, L., Ruotolo, F., Senese, V.P., Ruggiero, G., Masullo, M. and Alekseeva, N., **2012**. Multisensory assessment of acoustic comfort aboard metros: a virtual reality study. Applied Cognitive Psychology, 26(5), pp.757-767.
https://onlinelibrary.wiley.com/doi/pdf/10.1002/acp.2856
Less annoyance in AV, but performed worst in cognitive tasks.
Jeon 2018: "In some previous studies, the methods of evaluating actual noise were investigated by implementing a sound environment inside a train carriage via a head-mounted display (HMD), which is a commonly used device for visualizing virtual spaces. However, the evaluation results showed that the annoyance level of the noise inside the train tended to be underestimated [22]."
"may modulate loudness judgments. For example, colours such as red or pink seem to increase perceived loudness, whereas grey or pale green seem to decrease loudness (Menzel, Fastl, Graf, & Hellbrück, 2008).

Consistently, Parizet and Koehl (2011) found that pictures of red trains caused an increase in perceived loudness as compared with pale green trains."
Differences with levels above 80 dB SPL.



	SPL	N
Metros	dB(A)	Sone
Berlin S3	65.7	17.65
Paris L1	71.9	25.1
Turin L1	74.0	33.25
Berlin U2	75.85	33.5
Rome LA	76.25	27.1
Naples L1	80.6	37.95
New York MQ	81.2	47.95
Milan L1	82.15	46.3
Milan L3	86.2	55.35

Chau, C.K., Leung, T.M., Xu, J.M. and Tang, S.K., **2018**. Modelling noise annoyance responses to combined sound sources and views of sea, road traffic, and mountain greenery. The Journal of the Acoustical Society of America, 144(6), pp.3503-3513.
https://asa.scitation.org/doi/pdf/10.1121/1.5083833?class=pdf
Visual and audio annoyance. Video clips in a window panel of a living room containing sea, mountain greenery, and road. "246 participants were presented with 11 audio-visual stimuli. The findings revealed that views embracing mountain greenery close-by could enhance annoyance, which is contrary to other findings that greenery could always moderate noise annoyance."

Vallin, P.A., Marquis-Favre, C., Bleuse, J. and Gille, L.A., **2018**. Railway noise annoyance modeling: Accounting for noise sensitivity and different acoustical features. The Journal of the Acoustical Society of America, 144(6), pp.3381-3390.
Railway and tramway noise annoyance predictions. "Annoyance may be predicted using noise-exposure relationships with mean energy-based index."

ISO/TS 15666. Acoustics - Assessment of noise annoyance by means of social and socio-acoustic surveys. International Organization for Standardization: Genf; **2003** (2)
https://webstore.ansi.org/Standards/ISO/ISOTS156662003
Sun, K. 2018: "intercultural tested instrument to measure noise annoyance are already available. [ISO 15666]"
Preis et al. 2016: "the ICBEN scale (0–10) was used to rate the presented stimuli."
NoiseQ questionnaire to measure noise sensitivity. Where to find it? And the ICBEN scale?

Schutte, M., Marks, A., Wenning, E. and Griefahn, B., **2007**. The development of the noise sensitivity questionnaire. Noise and Health, 9(34), p.15.
http://www.noiseandhealth.org/article.asp?issn=1463-1741;year=2007;volume=9;issue=34;spage=15;epage=24;aulast=Schutte
NoiseQ noise sensitivity questionnaire. But I still don't find the questionnaire.
This is the questionnaire, but it doesn't tell how to calculate it:
http://www.noiseandhealth.org/viewimage.asp?img=NoiseHealth_2007_9_34_15_34700_10.jpg

Preis, A., Hafke-Dyx, H., Szychowska, M., Kocinski, J. and Felcyn, J., **2016**. Audio-visual interaction of environmental noise. Noise Control Engineering Journal, 64(1), pp.34-43.
https://scholar.google.de/scholar?hl=en&as_sdt=0%2C5&q=Audio-visual+interaction+of+environmental+noise&btnG=
Sun, K. 2018: "the complexity of the audiovisual interaction: in a lab experiment, adding visual information to a listening experiment tended to reduce annoyance if the sound source was believed to have a positive influence, while annoyance increased for mechanical sound sources."
"In this study the results of psychoacoustic experiments will be compared with objective noise characteristics such as LAeq,T, PA, N5, S, F, and R."
"In contrast, Viollon et al.11 argued that the common statement “the more urban visual setting, the more negative sound ratings” depends on the type of sound source"
Annoyance is not always reduced! Loudness N5 (sone non-stationary loudness model) is very correlated.

*Table 4—The percentage of explained variance (ω) calculated for sound characteristics in relation to the noise annoyance ratings of audio only or compatible audio-visual stimuli presentation (Pearson correlation test). Significant correlations (with $p < .05$) were marked with *.*

Sound characteristics versus ICBEN ratings	Explained variance ω					
	Urban (Airplane, Road, Train)					
Number of stimuli	For all stimuli 15		For AP. RD. TR 9		For SEA and FN 6	
ICBEN	Audio	Audio-video	Audio	Audio-video	Audio	Audio-video
N5	0.55*	0.3*	0.88*	0.86*	0.47	0.36
F	0.08	0.23	0.05	0.18	0.01	0.00
R	0.45*	0.20	0.71*	0.89*	0.85*	0.74*
S	0.00	0.09	0.17	0.49	0.46	0.58
PA	0.56*	0.29*	0.85*	0.86*	0.59	0.47

J.M. Fields, R.G. De Jong, T. Gjestland, I.H. Flindell, R.F.S. Job, S. Kurra, P.V. Lercher, T. Yano, R. Guski, U. Felscher-Suhr and R. Schuemer, “Standardized noise reaction 11 questions for community noise surveys: Research and a recommendation”, J. Sound Vibr., 242(4), 641–679, (2001).
Preis et al. 2016: "the ICBEN scale (0–10) was used to rate the presented stimuli." Also citing the ISO 15666

Marquis-Favre, C., Premat, E. and Aubrée, D., **2005**. Noise and its effects–a review on qualitative aspects of sound. Part II: Noise and annoyance. Acta Acustica united with Acustica, 91(4), pp.626-642.
Preis et al. 2016: "the most popular approach defines noise annoyance in terms of a single physical variable, i.e. sound pressure level, which is easily identified and measured both in field or laboratory conditions (for recent reviews see Marquis-Favre et al.15). Examples of such noise ratings are, among others, energy-based ratings such as LAeq,T, LDEN, and SEL"

Skovenborg, E. and Nielsen, S.H., **2004**, October. Evaluation of different loudness models with music and speech material. In Audio Engineering Society Convention 117. Audio Engineering Society.
Provided by Dirk. Signals containing strong tonal or impulsive components are generally considered to be more annoying than more random signals of the same level. A Danish guideline for environmental noise [58] therefore indicates that 5 dB(A) should be added in the case of clearly audible tonal components or clearly audible impulsive components in the noise.

Wu, Y.H. and Bentler, R.A., **2010**. Impact of visual cues on directional benefit and preference: Part II—field tests. Ear and hearing, 31(1), pp.35-46.

https://www.researchgate.net/publication/26830917_Impact_of_Visual_Cues_on_Directional_Benefit_and_Preference_Part_II-Field_Tests
4 weeks wearing hearing aids in the field. They selected two different modes (DIRectional and OMNIdirectional) and wrote it down. For two weeks they kept the eyes closed and for two other weeks they were asked to close the eyes and rate the best performance algorithm. They use "AV version of the Connected Speech Test (CST; Cox et al.1987, 1988, 1989)" to evaluate speech intelligibility and "Hearing In Noise Test (HINT; Nilsson et al. 1994)".
"the finding that **AV laboratory testing predicted field outcomes more accurately than did auditory-onl**y testing demonstrates the role of visual cues in the real world and emphasizes the necessity of using AV testing in the laboratory to evaluate DIR microphone hearing aids."

Raggam, R.B., Cik, M., Höldrich, R.R., Fallast, K., Gallasch, E., Fend, M., Lackner, A. and Marth, E., **2007**. Personal noise ranking of road traffic: Subjective estimation versus physiological parameters under laboratory conditions. International journal of hygiene and environmental health, 210(2), pp.97-105.
https://scholar.google.es/scholar?hl=en&as_sdt=0%2C5&q=Personal+noise+ranking+of+road+traffic&btnG=
Heart rate is related to very annoying sounds!
audio analysis software ArtemiS of Head Acoustics GmbH. the Psychoacoustics Module of the ArtemiS Analyses System (HEAD acoustics GmbH)

Cik, M., Fallast, K. and Höldrich, R.R., **2008**. Traffic Noise Annoyance on Roads (TNAR). Institute of Highway Engineering and Transport Planning, Graz Technology of University. Austria.
https://pdfs.semanticscholar.org/2204/c0025313e7a9ed83855360bcc7f49af41535.pdf
From the same group as Raggam 2007, but they are not cited (only 2). They create an annoyance measure but I didn't see any formulas.

CONCLUSIONS

· Greenery not always reduces annoyance (Chau 2018).
· Noise annoyance formulas (Preis et al. 2016). Something about noise type is mentioned in Sun et al. 2018.
· Visible sound sources → lower annoyance (Sun et al. 2018) and lower loudness (Aylor and Marks 1976).
· Annoyance is not always reduced with vision → more urban/tech vision equals to more annoyance (Preis et al. 2016)
· Annoyance formulas: energy-based: L(Aeq,T), L(DEN), and SEL; multicomponent: PA (au)(Psychoacoustic annoyance), N5 (sone), S (acum), F (vacil), and R (asper). They are based on the book from Fastl and Zwicker.

How to compute the annoyance and loudness?

The NI Sound and Vibration Measurement Suite provides LabVIEW VIs ISO 532B stationary loudness, time-varying loudness, Aures roughness, Aures sharpness, Aures tonality, and fluctuation strength.

The Psychoacoustics Module of the ArtemiS Analyses System (HEAD acoustics GmbH)
https://www.head-acoustics.com/downloads/eng/artemis/D5012_ArtemiS_suite_Psychoacoustics_Module_e.pdf

Salford Sound Quality

https://www.salford.ac.uk/research/sirc/research-groups/acoustics/psychoacoustics/sound-quality-making-products-sound-better/accordion/sound-quality-testing/matlab-codes

GENESIS S.A - Loudness
http://www.genesis.fr

Cabrera, D., **1999**, November. PSYSOUND: A computer program for psychoacoustical analysis. In Proceedings of the Australian Acoustical Society Conference (Vol. 24, pp. 47-54).
http://psysound.org/
Could not make it work… Only roughness

LOUDNESS SCALES

Heeren, W., Hohmann, V., Appell, J.E. and Verhey, J.L., **2013**. Relation between loudness in categorical units and loudness in phons and sones. The Journal of the Acoustical Society of America, 133(4), pp.EL314-EL319. https://asa.scitation.org/doi/abs/10.1121/1.4795217
From Dirk Oetting. Relationship between CU, phones and sones.
ISO 226, 2003; ANSI S3.4, 2007; DIN 45631, 1991; DIN 45631/A1, 2010; ISO 16832, 2006
"To relate loudness in CUs to loudness in sones, the transformation from phons to sones as tabulated in ANSI S3.4 (2007) was used."
"loudness categories were linearly transformed to numerical values (CUs) from 0 (inaudible) to 50 (too loud)."

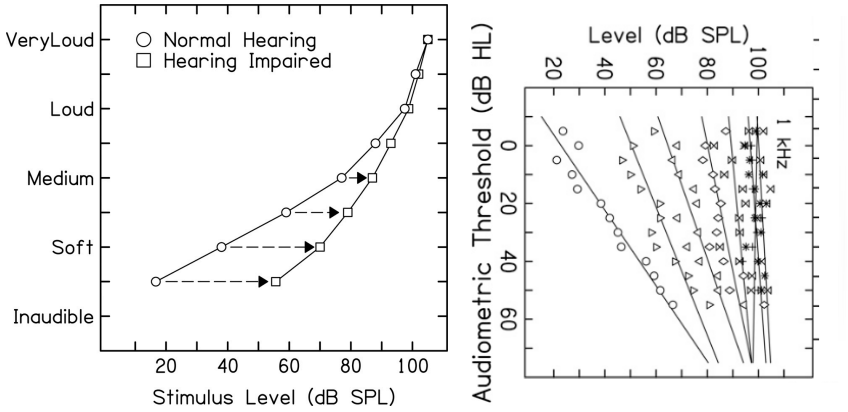
From References provided by Dirk Oetting
Scheuern, J.. **2014**. ISO 532 – Living and working with alternative loudness standards. Internoise
ISO 532-1 and ISO 532-2. Two standards, both valid. Moore/Glasberg and Zwicker. Seems that there is a lot of names with the ISOs: ISO 226, DIN 45631, ANSI S3.4. Other things in the article: ISO TC 43 “Acoustics” WG 9 “Method for calculating loudness level".

ISO (**2006**). 16832:2006. “Acoustics—Loudness scaling by means of categories” (International Organization for Standardization, Geneva, Switzerland).
https://cdn.standards.iteh.ai/samples/32442/13db0e73b8714d13832343ede0909a91/ISO-16832-2006.pdf
Heeren: "provides an easy and fast procedure for measuring loudness over the subject’s whole dynamic range. Using this procedure the loudness is determined on a scale with named categories such as inaudible, very soft, soft, medium loud, loud, very loud, and too loud or extremely loud."
Heeren: "non-adaptive loudness scaling procedure according to ISO 16832 (2006). This procedure included two stages. In the first stage, the dynamic range of the listeners was estimated by presenting a sequence of stimuli with increasing level. The listener’s task was to press a response button as soon as the stimulus was audible and press it a second time when it was too loud."
Heeren: "In order to reduce context effects (the tendency of some listeners to rate the current stimulus relative to the previous stimulus), the stimuli were presented in pseudo-random order where the maximum difference of successive presentation levels was smaller than half of the dynamic range of the sequence, in agreement with ISO 16832 (2006)."

Brand, T., and Hohmann, V. (**2002**). “An adaptive procedure for categorical loudness scaling,” J. Acoust. Soc. Am. 112(4), 1597–1604.
Heeren: "A model loudness function (i.e., CU vs level) was fitted to the individual data for each frequency as described by Brand and Hohmann (2002). The function consists of two linear parts with independent slope values and a transition region which was smoothed using a Bezier fit. The function was fitted to the individual data using a modified least-square fit according to Brand and Hohmann (2002). Using the fitted functions, the levels corresponding to given loudness values in CUs were derived from the individual loudness functions for each frequency."

Al-Salim, S.C., Kopun, J.G., Neely, S.T., Jesteadt, W., Stiegemann, B. and Gorga, M.P., **2010**. Reliability of categorical loudness scaling and its relation to threshold. Ear and hearing, 31(4), p.567.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3376894/
Good introduction and discussion, good for dissertation and knowledge.

Average dB SPL for each loudness category. They test the mean and the median of the dB SPL, also removing outliers.
Nice figure about what is the problem with loudness and hearing impaired participants. Loudness recruitment "Loudness recruitment, observed in patients with hearing loss, has been attributed to the spread of excitation as intensity increases (Scharf, 1978), to the loss of compressive nonlinearity, which typically accompanies hearing loss of cochlear origin (Moore, 2003), or perhaps to a combination of these and/or other effects determined at more central levels of the auditory system (Heinz et al., 2005; Cai et al., 2009). "



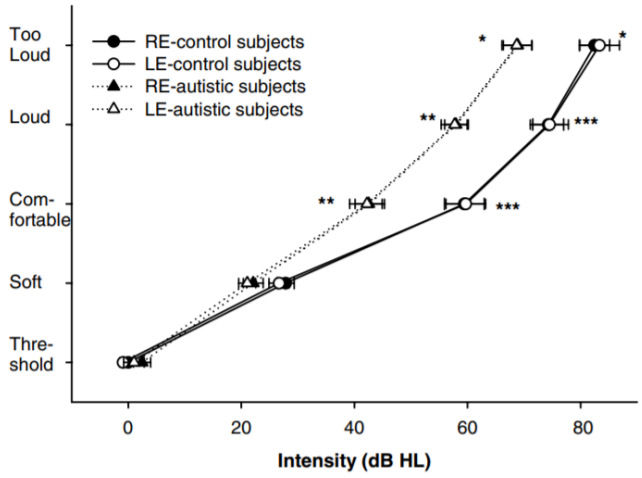
"(Ellis and Wynne, 1999). Beattie et al. (1997) reported test-retest values within a given category of ± 12 dB 95% of the time in a single session"
"Regardless of whether mean or median levels were used, reported standard deviations were in the range of 1.4 to 9 dB with test-retest differences of less than 10 dB for adults and greater than 10 dB for children (Ellis and Wynne, 1999). Ricketts and Bentler (1996)"
"the mean SPL assigned to a category in session 1 and the mean SPL for the same category in session 2, were less than 1 dB at all three test frequencies and the standard deviations of the average stimulus-level differences ranged from 6.6 dB to 7.8 dB."

Rasetshwane, D.M., Trevino, A.C., Gombert, J.N., Liebig-Trehearn, L., Kopun, J.G., Jesteadt, W., Neely, S.T. and Gorga, M.P., **2015**. Categorical loudness scaling and equal-loudness contours in listeners with normal hearing and hearing loss. The Journal of the Acoustical Society of America, 137(4), pp.1899-1913.
https://doi.org/10.1121/1.4916605
http://audres.org/cel/cls/Rasetshwane2015JASA.pdf
"**loudness recruitment**, a phenomenon in which the rate of loudness growth with sound intensity is more rapid than normal (e.g., Steinberg and Gardner, 1937; Stevens and Guirao, 1967; Scharf, 1978; Hellman and Meiselman, 1993)"
"CLS (1) is easy to perform and relates to a listener's experience and informal descriptions of his or her loudness percepts, (2) requires minimal training, (3) is reliable, and (4) requires an amount of time that might be acceptable clinically (Allen et al., 1990; Brand and Hohmann, 2002; Al-Salim et al., 2010; Oetting et al., 2014)."
"Despite its advantages, CLS has not been widely accepted for clinical applications, in part, because of concerns related to the reliability of the measurements (e.g., Elberling, 1999), its use of arbitrary units (CUs) that do not relate to standard units of loudness or loudness level, its use of bounded categories and its underestimation of the slope of loudness growth (Hellman, 1999). Another criticism is that CLS lacks the properties of a ratio scale (Stevens, 1946), i.e., it does not quantify how much louder one sound is compared to another; it only provides information regarding the ranking of sounds from softest to loudest. While previous

studies have demonstrated the reliability of CLS measurements (e.g., Robinson and Gatehouse, 1996; Cox et al., 1997; Rasmussen et al., 1998; Al-Salim et al., 2010), these studies focused on group data, instead of data from individual participants. And last, although several researchers have considered using CLS for fitting hearing aids, it remains unclear how these data should be incorporated into the fitting and, in cases where it is used, only limited benefit has been demonstrated (e.g., Cox, 1995; Ricketts, 1996). Consequently, further evaluation of CLS measurements is required if they are going to be accepted for clinical use."

Ellis, M.R. and Wynne, M.K., **1999**. Measurements of loudness growth in 1/2-octave bands for children and adults with normal hearing. American Journal of Audiology, Rockville Vol 8, 1
https://www.proquest.com/docview/204378279/fulltextPDF/A712625A92244B18PQ/1?accountid=10336
Average between CU for a group type and condition.

Khalfa, S., Bruneau, N., Rogé, B., Georgieff, N., Veuillet, E., Adrien, J.L., Barthélémy, C. and Collet, L., **2004**. Increased perception of loudness in autism. Hearing research, 198(1-2), pp.87-92.
https://www.sciencedirect.com/science/article/pii/S0378595504002308?casa_token=XeDvmHAAvqwAAAAA:6pyXEm_LXCMafEUz6CO-v1seeVtE5W-af9-r205Q7Yx-CyoOfJl_kjiIo3A6wWqHhuTBrvw0Tg#fig1
Average of dB SPL for each loudness category.



SIMULATOR SICKNESS

Vovk, A., Wild, F., Guest, W. and Kuula, T., **2018**, April. Simulator Sickness in Augmented Reality Training Using the Microsoft HoloLens. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (p. 209). ACM.
https://dl.acm.org/citation.cfm?id=3173783
7.03 to 10.08 total score for SSQ is declared as negligible symptoms.

Singla A., S. Fremerey, W. Robitza, and A. Raake. **2017**. "Measuring and Comparing QoE and Simulator Sickness of Omnidirectional Videos in Different Head Mounted Displays." In 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), 1–6. doi:10.1109/QoMEX.2017.7965658.
https://ieeexplore.ieee.org/abstract/document/7965658/?part=1
"In order to watch the 360◦ video using the Whirligig player, projection type and Field of View (FOV) has to be set to Barrel and 360 respectively. We used the Whirligig player (version 3.89) because of its compatibility with both HMDs and the fact that it can be accessed via command line."

HEAD TRACKING WITH HMD HTC VIVE

Niehorster, D.C., Li, L. and Lappe, M., **2017**. The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research. i-Perception, 8(3), p.2041669517708205.
http://journals.sagepub.com/doi/abs/10.1177/2041669517708205

HRTF DISTORTIONS HMD

Genovese, A., Zalles, G., Reardon, G. and Roginska, A., **2018**, August. Acoustic perturbations in HRTFs measured on Mixed Reality Headsets. In Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality. Audio Engineering Society.
https://s18798.pcdn.co/immersiveaudiogroup/wp-content/uploads/sites/7671/2018/09/Acoustical_distortions_from_Augment_Reality_devices.pdf
"may affect localization at specific locations, potentially leading to lateralization effects or vertical localization errors in mixed reality auditory settings"
"Neither the ITDs nor the ITD frequency range were significantly affected by HMD devices."
"The HTC Vive, GearVR and Oculus Rift VR HMDs were also measured. Data to be released and discussed in a future publication"
"The quantitative observations are deemed to be nonnegligible, enough to motivate further perceptual study."

Gupta, R., Ranjan, R., He, J. and Woon-Seng, G., **2018**, August. Investigation of effect of VR/AR headgear on Head related transfer functions for natural listening. In Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality. Audio Engineering Society.
http://spacelab.spatialhearing.org/reprints/stecker2018avar.pdf
Timbre and spatial quality evaluation with HTC Vive, Rift and another one. Recorded sounds with a KEMAR and evaluated with subjects with MUSHRA.

Ahrens, A., Lund, K.D., Marschall, M. and Dau, T., **2018**. Sound source localization with varying amount of visual information in virtual reality. bioRxiv, p.489484.

https://www.biorxiv.org/content/biorxiv/early/2018/12/07/489484.full.pdf
Localization with HTC Vive, Blindfolded and with cues. HMD has an effect, making the localization accuracy worse.

LISTENING EFFORT

Speech recognition accuracy is not an indicator of listening effort (Fraser 2010, Winn 2015). Common measures are accuracy (on primary and/or secondary task), reaction time and subjective questionnaires.

Pichora-Fuller, M. Kathleen; Kramer, Sophia E.; Eckert, Mark A.; Edwards, Brent; Hornsby, Benjamin W.Y.; Humes, Larry E.; Lemke, Ulrike; Lunner, Thomas; Matthen, Mohan; Mackersie, Carol L.; Naylor, Graham; Phillips, Natalie A.; Richter, Michael; Rudner, Mary; Sommers, Mitchell S.; Tremblay, Kelly L.; Wingfield, Arthur. **2016**. Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL)
http://journals.lww.com/ear-hearing/Abstract/2016/07001/Hearing_Impairment_and_Cognitive_Energy___The.2.aspx
Using a three-dimensional illustration, we highlight how listening effort depends not only on hearing difficulties and task demands but also on the listener's motivation to expend mental effort in the challenging situations of everyday life.
Seems like an extensive paper written by a lot of people.

Lunner, T., Rudner, M., Rosenbom, T., Ågren, J. and Ng, E.H.N., **2016**. Using speech recall in hearing aid fitting and outcome evaluation under ecological test conditions. Ear and hearing, 37, pp.145S-154S.
https://www.eriksholm.com/news/2018/listening-effort-workshop
https://www.eriksholm.com/-/media/eriksholm/main/files/research-learn-more/other/lunner-et-al-2016-eh.pdf?la=en&hash=7A878272758380DD0029737D1926B7823F2BBB54
SWIR-test (sentence-final word identification and recall test) for measuring listening effort

Dimitrijevic, A., Smith, M.L., Kadis, D.S. and Moore, D.R., 2019. Neural indices of listening effort in noisy environments. Scientific reports, 9(1), pp.1-10.
https://www.nature.com/articles/s41598-019-47643-1
From Valerie, some interesting images about brain areas and listening effort. Published in Nature?

LISTENING EFFORT WITH VISUAL CUES

With the same levels of SNR, AV improves accuracy and perceived listening effort, but the RT is not improved.
With the same speech recognition level (SRT), AV takes more RT, is perceived more effortful and decreases accuracy in the secondary task.
Thus, AV only helps the speech recognition level but doesn't impede the increase of listening effort (accuracy on secondary task, RT and subjective listening effort).

Picou et al. could not find any significant changes in the RT in AV-AO experiments with the same speech recognition level and proclaimed inconsistent results with Fraser (2010). Nevertheless it is mentioned in their papers (2011) that AV conditions might take more time due to audiovisual integration.

Brown, V.A. and Strand, J.F., **2019**. About face: Seeing the talker improves spoken word recognition but increases cognitive effort.
https://osf.io/86zdp/
Introduction about contradictory findings. Email the author. Four conditions. Easy-Hard, AV, AO. Vibrotactile stimuli.

Annelies Devesse, Alexander Dudek, Astrid van Wieringen & Jan Wouters. **2017.** The AVATAR-approach: How real-life listening affects speech intelligibility and listening effort
Financed by OTICON.
Realistic 3D environments over a big 2D flat projection. Several virtual characters and dual-task paradigm.
Speech in noise.

Fraser, S., Gagné, J. P., Alepins, M., et al. (**2010**). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: the effects of providing visual speech cues. J Speech Lang Hear Res, 53, 18–33.
http://jslhr.pubs.asha.org/article.aspx?articleid=1782571
Picou E: "the primary task, which were closed-set sentence recognition".
Picou E: "Visual cues did not change listening effort when word recognition was matched in AO and AV conditions. These results are consistent with those reported by Picou et al. (2011) and Picou et al. (2013), but are inconsistent with the results reported by Fraser et al., (2010) and Mishra et al. (2013)" → Picou measures RT and Fraser Questionnaires?

Exp1: same noise, AV vs AO. Exp2: same SI levels.
In Experiment 1, the AV modality speech recognition was rated as less effortful, and accuracy scores were higher than A only. In Experiment 2, reaction times were slower, tactile task performance was poorer, and listening effort increased, in the AV versus the A-only modality. At equivalent accuracy levels, the dual-task performance decrements (for both tasks) suggest that the noisier AV modality was more effortful than the A-only modality.

Reaction times were not significantly different between AO-AV but questionnaires on Listening Effort were in Experiment 1.

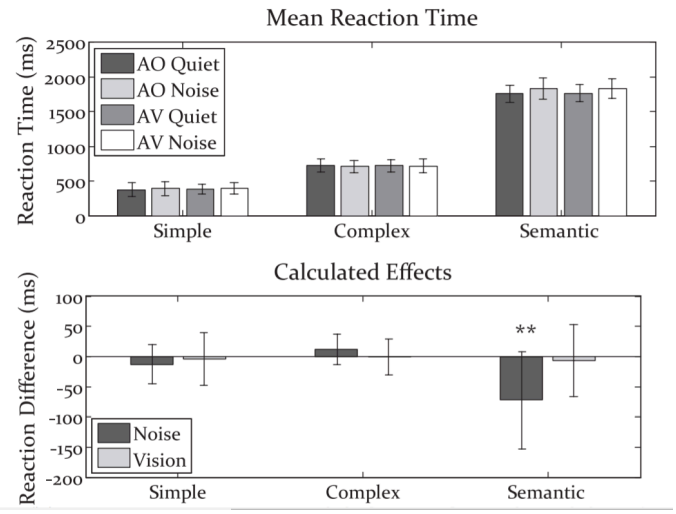
Picou, E.:
Always monosyllabic word recognition. Recorded video (television studio) as AV stimuli. Results differ from other studies (Fraser et al. 2010). Supported by her and others, audiovisual integration takes more effort (actually reaction times), although it improves recognition.

Picou, E. M., Gordon, J., & Ricketts, T. A. (**2016**). The Effects of Noise and Reverberation on Listening Effort in Adults With Normal Hearing. Ear Hear, 37(1), 1-13.
Jana: "Dual study task. Secondary task is driving a car simulator. Concurrent stimuli with same modality, same stimuli"
YNH; N=18; 22-30 (24.8). Effect of noise and reverberation.
RTs not correlated with word recognition, neither in quiet nor in babble and at no reverb level;
The finding that reverberation did not affect listening effort, even when word recognition performance was degraded, is inconsistent with current models of listening effort. However, the results of this study are limited in generalizability to young listeners with normal hearing and to the signal to noise ratios, loudspeaker to listener distance, and reverberation times evaluated.

Picou, E. M., & Ricketts, T. A. (**2014**). The Effect of Changing the Secondary Task in Dual-Task Paradigms for Measuring Listening Effort. Ear Hear, 35(6), 611-622.

http://journals.lww.com/ear-hearing/Abstract/2014/11000/The_Effect_of_Changing_the_Secondary_Task_in.4.aspx
Three different dual-tasks.
Gagne 2017: "It is difficult to determine whether the secondary task that consisted of the word-category recognition task was more sensitive to listening effort because it involved processing linguis- tic information (just as the primary task did) or because it was cognitively more demanding in other respects."
Visual cues did not change listening effort when word recognition was matched in AO and AV conditions. These results are consistent with those reported by Picou et al. (2011) and Picou et al. (2013), but are inconsistent with the results reported by Fraser et al., (2010) and Mishra et al. (2013). The results of the current experiments suggest that neither the complex paradigm nor semantic paradigm were sensitive to the effects of visual cues. While the reasons underlying the discrepant reports in the literature remain unclear, differences across studies may not be explained by secondary task complexity or depth of processing. The Fraser et al. (2010) and Picou et al. (2013) studies also varied in terms of the primary task, which were closed-set sentence recognition and open-set monosyllable word recognition, respectively.

Jana: "Dual study task. Secondary task is RT to visual probe/distractor and Semantic judgment task. Concurrent stimuli with same modality, same stimuli // Concurrent stimuli with different modalities.".
YNH; N=17, 21-24. HI; N=17, 23-73 (60.1) years; sensorineural HL, unaided testing.
Red/white rectangle, choice of response and world-classification.
AV performance benefit in quiet, no diff in babble
Monosyllable word recognition



Picou, E. M., Ricketts, T. A., & Hornsby, B. W. (**2013**). How hearing aids, background noise, and visual cues influence objective listening effort. Ear Hear, 34(5), e52-64.
Jana: "Dual study task. Secondary task is RT to visual probe/distractor. Concurrent stimuli with different modalities.Effects assessed: modality, amplification. HI with HA Phonak Savia 211 BTEs, N=27, 49-89 years. "
Similar experiment from 2011, with some different metrics. AV had greater effect on quiet-babble differences, but only better in quiet. Overall RT effects not significant.
Short memory benefited more from visual cues (only RT)
AV performance benefit in quiet, no diff in babble
"Probably" monosyllable word recognition

With regard to visual cues, while there was no effect on average of providing visual cues, there were some listener characteristics that were related to changes in listening effort with vision. Although these relationships are exploratory, they do suggest that these inherent listener characteristics like working memory capacity, verbal processing speed, and lipreading ability may influence susceptibility to changes in listening effort and thus warrant further study.

Picou, E. M., Ricketts, T. A., & Hornsby, B. W. (**2011**). Visual cues and listening effort: individual variability. J Speech Lang Hear Res, 54(5), 1416-1430.
http://jslhr.pubs.asha.org/article.aspx?articleid=1782642
http://jslhr.pubs.asha.org/epdf.aspx?doi=10.1044/1092-4388(2011/10-0154)
No significant results between AV and AO at the same SI level. However, reaction times for these tasks were slower in AV conditions than in AO conditions, suggesting that although visual cues helped improve performance, processing the additional visual information required additional cognitive resources.
Monosyllable word recognition

Moradi2013: "Our results are not consistent with those of Picou et al. (2011), which showed that low working memory capacity was associated with relatively effortful audiovisual identification of stimuli in noise."

Jana: "Dual study task. Secondary task is Recall of words represented in the primary task or before the primary task. . Concurrent stimuli with same modality same stimuli. NH (normal hearing), N=20, 19-44 years. Effects assessed: modality, masker type, effect of WMC and lipreading ability on A-only vs AV effort"
Primary task: Recognition of monosyllabic words in quiet or 4-talker bable; sets of 5 words; individual SNR for 75%-correct; conds: A-only vs AV; outcome: %-correct
Secondary task: recall of presented words (sets of 5 words); outcome %-correct
Measures: subjective degree of effort (11 point likert). Lipreading skills, Revised Shortened Utley Sentence Lipreading Test (ReSULT), visual-only sentence recognition. WM capacity, automated operation span task (AOSPAN), solving of simple equations, memorization of letters
Results: performance equal for A-only and AV. better recall for positions 4 and 5 than 1-3.

Objective and subjective results indicated that listening effort increased in the presence of noise, but on average the addition of visual cues did not significantly affect the magnitude of listening effort.

"Although there was substantial individual variability, on average participants who were better lipreaders or had larger WMCs demonstrated reduced listening effort in noise in AV conditions. The results support the hypothesis that integrating auditory and visual cues requires cognitive resources in some participants."

Anderson-Gosselin, P. A., & Gagné, J. P. (**2011a**). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. Int JAudiol, 50(11), 786-792.

Wu, Y. H., Aksan, N., Rizzo, M., Stangl, E., Zhang, X., & Bentler, R. (**2014**). Measuring Listening Effort: Driving Simulator Versus Simple Dual-Task Paradigm. Ear Hear, 35(6), 623-632.

Winn, Matthew B., Jan R. Edwards, and Ruth Y. Litovsky. “The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation.” Ear and hearing 36.4 (**2015**): e153–e165. PMC. Web. 9 May 2017. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4478109/
Seems like a pretty straight forward finding of pupil dilation with degraded speech (vocoded, cochlear implant simulation).

McGarrigle, R., Munro K.J., Dawes, P., Stewart A. J., Moore D. R., Barry, J. G., Amitay, S.. **2014**. Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. In International Journal of Audiology 53, 433-445
Selective gain: filtering out irrelevant information (Kerlin et al 2010). With hearing loss - more effort to listen in noisy conditions.
Fatigue: However, despite the intuitive link between the effort involved in real-time listening and the (more long term) fatigue that emerges as a consequence of that effort, there is currently little empirical support for this connection within hearing research.
Self-reported listening effort (general, everyday life situation): An example of the questionnaire method can be found in the multi-dimensional speech, spatial, and qualities (SSQ) hearing scale (Gatehouse and Noble, 2004) which measures the extent of listening difficulties experienced in various real-world settings, e.g. with and without access to visual cues (49 items). In Dawes et al (2014), the authors constructed a ‘ listening effort ’ subscale from the effort-related questions in the SSQ hearing scale in order to examine changes in listening effort subsequent to acclimatization to hearing aids.
Self-reported judgements of effort (during experiments): In one large-scale study, participants were asked to indicate (on a 100-point scale) how ‘ effortful ’ they found each particular trial in an auditory profile test battery (van Esch et al, 2013). Such self-report scales are often used in conjunction with physiological techniques (Mackersie & Cones, 2011; Zekveld et al, 2011; Koelewijn et al, 2012). *some individuals may use task performance accuracy (or task difficulty) rather than mental exertion to inform their self-reported effort rating. Therefore, changes in the mental exertion required to understand an auditory message may go undetected using self-report measures alone.*
Self-reported fatigue: *For example, a subject in a relatively positive state of mind might underestimate the extent of their fatigue and vice versa (DeLuca, 2005).*
Behavioral responses
Single task: Piquado et al (2012) showed that speech comprehension difficulties in hearing- impaired individuals could be alleviated by allowing the participant to control the rate of speech.

Response time: Despite optimal intelligibility, response times were found to be significantly slower in the more challenging signal-to-noise ratios (SNRs). The authors interpret this as reflecting the increased listening effort required to understand speech in difficult acoustic conditions (Houben et al, 2013). This behavioural effect may be interpreted as slowed phonological processing and/or lexical access, the impact of which is thought to be modulated by an individual’s working-memory capacity (Rönnberg et al, 2008). However, the relationship between the effort required to understand an auditory stimulus, and the timing of a response to a stimulus, remains unclear. For example, it is equally plausible that an individual may respond more quickly to a stimulus as a result of more focused attention because of the need to exert more cognitive effort to its processing.
Multi-task
For measurement of listening effort, a speech recognition task (in varying SNRs) is often used as the primary task. Secondary tasks may include, for example, a memory (Howard et al, 2002), tactile recognition (Anderson Gosselin & Gagné , 2011), or visual recognition (Hicks & Tharpe, 2002) task. Appears to have good face validity with regard to speech processing in realistic environments. However, it remains unclear whether

secondary task performance actually constitutes an objective index of listening effort (i.e. ‘ mental exertion ’) per se.

Behavioral measures of fatigue

The ‘ slowing ’ or decline of cognitive functions following concerted and/or prolonged mental effort (DeLuca, 2005). Hornsby ’ s (2013) finding provides the first empirical link in the assumed relationship between repeated incidences of effortful listening and subsequent cognitive fatigue. Fatigue was indexed as the relative increase in visual task response time (i.e. response slowing) over the duration of each one-hour aided and unaided trial block.

Physiological measures of listening effort

Functional Magnetic Resonance Imaging (fMRI)

Consequences of neuronal activity and the resulting changes in blood oxygenation level. Increased activity in the left inferior frontal gyrus was found when participants were required to attend to degraded speech, as compared to clear speech. This was interpreted as reflecting the compensatory effort required to perform challenging listening tasks. Further research with hearing-impaired individuals in which realistic listening environments are simulated may elucidate whether this kind of increase in frontal activity is indicative of listening effort.

Another limitation on experimental design is head motion, which can lead to artificial intensity changes of the fMRI signal.

Electroencephalography (EEG)

Cortical EEG involves measuring electric potential fluctuations associated with neural activity through a series of electrodes placed directly on the scalp. EEG activity is generally categorized in terms of the frequency ‘ band ’ to which particular oscillations (regular waveform fluctuations) belong. Earlier peaks and increased amplitude of the N1 ERP component (a negative polarity potential occurring 100 ms post stimulus-onset) when processing more- degraded speech, compared to less-degraded speech. However, it remains unclear whether these markers can tell us about the kind of ‘ mental exertion ’ likely to give rise to the chronic experience of fatigue and tiredness in hearing-impaired individuals. In order to make this connection, it may be worthwhile to look at changes in spectral or temporal characteristics over the time period during which fatigue may be induced (in the order of minutes/hours).

Pupillometry

The changing size of the eye's pupil is controlled by muscle activity in the iris. Pupil size fluctuates as an adaptive response to changes in the environment (e.g. the ‘ light reflex ’). However, fluctuations in pupil size may also be elicited by changes in mental task load (Kahneman, 1973). Older adults have been shown to have relatively small absolute pupil sizes (Zekveld et al, 2011) which may influence the extent to which a significant change in pupil size can be recorded. In contrast, children show larger absolute pupil sizes. Overall, in well-controlled experimental settings, relative changes in pupil size may reflect systematic changes in ‘ mental exertion ’ not otherwise captured using behavioural measures alone.

Skin conductance

Amount of moisture produced by these sweat glands are influenced by parasympathetic and sympathetic activity reflecting changes in arousal. accuracy was similar across conditions, but significantly higher skin conductance responses were found for medium- and high-demand listening conditions. Changes in skin conductance also appeared to be reasonably consistent at the individual level. Sixty percent of participants showed significant skin conductance changes from low- to medium- and high-demand conditions (Mackersie & Cones, 2011). The authors found a correlation (r 0.67) between changes in skin conductance and self-reported listening effort across task demand (Mackersie & Cones, 2011). Other physiological measures were included in this study (skin temperature, frontalis muscle activity, and heart rate), with only muscle activity showing

significant task- related changes. However, these changes in muscle activity were not found to be as consistent across individual participants as skin conductance.

Physiological measures of fatigue

To our knowledge, only one attempt at measuring listening-related fatigue using physiological techniques has been made. Hicks and Tharpe (2002) examined differences in cortisol levels between hearing-impaired and normal-hearing children for potential markers of fatigue. Salivary samples were taken at the beginning and end of the school day and no significant differences in average cortisol levels were found between hearing-impaired and normal-hearing children.

Jing Xia, Sridhar Kalluri, Christophe Micheyl, and Ervin Hafter. **2017**. Continued search for better prediction of aided speech understanding in multi-talker environments. In ASA
http://asa.scitation.org/doi/10.1121/1.5008498#fragmentNav_0

Phonetic vs Semantic conditions. With hearing aids, phonetic is improved, but semantics presents a large individual variability (some improve some other not). This effect cannot be explained. Improving SNR had larger impact on phonetics than semantics.

Stimuli is stories. Competing speaker speaking one of the stories, separated 15º from target speaker. Another experiment with six-talker speech babble noise at -7.5º and 7.5º. "Traditional speech recognition tasks have not been especially effective in predicting the benefits of hearing technology in real-world listening. We suggest that asking about meaning versus speech reception should be more informative about hearing-impaired listeners' experience with hearing aids."

AGING AND HEARING LOSS

Samira Anderson. **2018**. Hearing and Aging Effects on Speech Understanding: Challenges and Solutions. Acoustics Today Acoustical Society of America, vol 14((4)
https://acousticstoday.org/wp-content/uploads/2018/12/Hearing-and-Aging-Effects-on-Speech-Understanding-Samira-Anderson.pdf?utm_source=ASA&utm_medium=email&utm_campaign=10149953_AT%202018%20winter%20issue&dm_i=1ZJV,61JR5,MVZI1S,NPBB6,1
Talks about the ineffectiveness of the audiogram, the hidden hearing loss and other current problems related with aging. "The Audiogram Does Not Predict Speech Understanding in Noise", "Hidden Hearing Loss", “Don’t Talk So Fast!”, “Why Are You Shouting?”, “Why Is Speech So Unclear?”, “Why Do I Still Have Trouble Understanding Speech with My Hearing Aids?”
The group has also done some research related to music.

Souza, P.E., Boike, K.T., Witherell, K. and Tremblay, K., **2007**. Prediction of speech recognition from audibility in older listeners with hearing loss: Effects of age, amplification, and background noise. Journal of the American Academy of Audiology, 18(1), pp.54-65.
https://www.researchgate.net/profile/Pamela_Souza/publication/6550847_Prediction_of_Speech_Recognition_from_Audibility_in_Older_Listeners_with_Hearing_Loss_Effects_of_Age_Amplification_and_Background_Noise/links/5689548508ae051f9af7677a.pdf
"performance worsened significantly with increasing age, even among listeners younger than 75 years"
"We found no interaction between age and type of noise. The small reduction in scores for amplitude-modulated compared to steady noise and lack of age interaction suggests that the substantial deficit seen with age in multitalker babble for previous studies was due to some effect not elicited here, such as informational masking."

Gordon-Salant, S., **2014**. Aging, hearing loss, and speech recognition: stop shouting, I can’t understand you. In Perspectives on auditory research (pp. 211-228). Springer, New York, NY.
They also talk about audiovisual speech integration, maybe interesting to give it a quick read.
"Another problem in successful use of hearing aids by older listeners is the combination of reduced spectral resolution, due to the sensorineural hearing loss, coupled with **reduced auditory temporal processing** associated with age. Souza and her colleagues (e.g., Souza, 2000) have confirmed that an age-related deficit in the use of temporal cues has an impact on older listeners’ ability to process temporal-envelope cues in speech with compression amplification and/or across channels"

Mishra S. , Lunner T. , Stenfelt S. , Rönnberg J. & Rudner M . **2013** . Seeing the talker's face supports executive processing of speech in steady state noise . Front Syst Neurosci , 7 : 96 . DOI: 10.3389/fnsys.2013.00096 .
McGarrigle et al.: Such interference, however, may be modulated by the presence of a visual aid (i.e. seeing the talker’s face), which has been found to reinforce the executive processing of speech in steady state noise (Mishra et al, 2013).

Rudner, M., Mishra, S., Stenfelt, S., Lunner, T., & Rönnberg, J. (**2016**). Seeing the Talker's Face Improves Free Recall of Speech for Young Adults With Normal Hearing but Not Older Adults With Hearing Loss. Journal of Speech, Language, and Hearing Research, 59(3), 590-599.

Tye-Murray, N., Sommers, M.S. and Spehar, B., **2007**. The effects of age and gender on lipreading abilities. Journal of the American Academy of Audiology, 18(10), pp.883-892.
https://www.ingentaconnect.com/content/aaa/jaaa/2007/00000018/00000010/art00007
"The results suggest that while **older adults do not lipread as well** as younger adults, the difference between older and younger participants was comparable across gender"

Tye-Murray, N., Sommers, M. S., & Spehar, B. (**2007**). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. Ear and hearing, 28(5), 656-668.
https://journals.lww.com/ear-hearing/Abstract/2007/09000/Audiovisual_Integration_and_Lipreading_Abilities.6.aspx
"The one exception to this finding was that the participants with hearing impairment performed significantly better than the participants with normal hearing on V identification of words. Measures of visual enhancement, auditory enhancement, and auditory-visual integration did not differ as a function of hearing status."
Jamaluddin 2016: "This indicates that performance of lipreading is highly individual and not related entire to age and hearing impairment." "hearing impairment is not a factor in a person's ability use both auditory and visual information in speech perception"

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., Hale, S. and Rose, N.S., **2008**. Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions. International Journal of Audiology, 47(sup2), pp.S31-S37.
https://www.tandfonline.com/doi/full/10.1080/14992020802301662
Older adults worst in speech recognition.
Jamaluddin 2016: "the performance of older adults deteriorated more compared to young adults when the viewing and listening conditions were unfavourable."
"For example, Sommers et al (2005) also found that older adults were less able to recognize audiovisual sentence materials than were young adults, and other investigators have shown that older adults are more affected by increases in background noise (Pederson et al, 1991; Plath, 1991; GordonSalant & Fitzgibbons, 1999)."

Sommers, M.S., Tye-Murray, N. and Spehar, B., **2005**. Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. Ear and hearing, 26(3), pp.263-275.
https://journals.lww.com/ear-hearing/Abstract/2005/06000/Auditory_Visual_Speech_Perception_and.3.aspx
Andre 2016: "found that older adults integrate consonants and words less well than younger adults, but do as well with sentence stimuli. suggest that hearing threshold matched older adults should perform as well as younger adults under real-world listening conditions"
→ I do not agree, the plots do not show this.

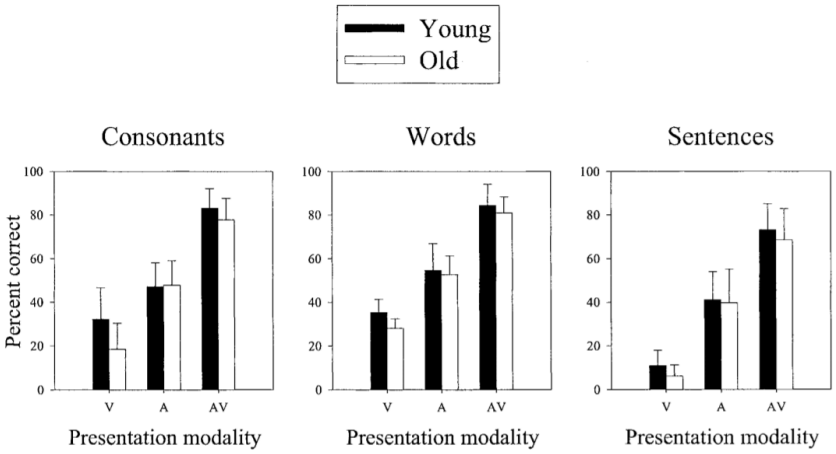


Fig. 1. Percent correct identification of consonants (left), words (middle), and sentences (right) for younger (filled bars) and older (open bars) adults. Presentation modality refers to visual only (V), auditory only (A), and auditory-visual (AV). Error bars indicate standard deviation.

Tillberg, I., Rönnberg, J., Svärd, I. and Ahlner, B., **1996**. Audio-visual Speechreading in a Group of Hearing Aid Users the Effects of Onset Age, Handicap Age, and Degree of Hearing Loss. Scandinavian Audiology, 25(4), pp.267-272.
https://www.tandfonline.com/doi/pdf/10.3109/01050399609074966
No audiovisual differences, but **early on-set** significantly better in visual-only.

Zhou, X., Innes-Brown, H. and McKay, C.M., **2019**. Audio-visual integration in cochlear implant listeners and the effect of age difference. The Journal of the Acoustical Society of America, 146(6), pp.4144-4154.
https://asa.scitation.org/doi/full/10.1121/1.5134783
CI and NH and elderly. CVC words and VCV (/aSa/). They use reaction time and accuracy as measure. No differences found: "experienced CI listeners showed no better AV integration than their similarly aged NH adults. Further, there was no significant difference in AV integration between the younger and older NH adults."

Füllgrabe, C. and Rosen, S., **2016**. On the (un) importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. Frontiers in psychology, 7, p.1268.
Moradi 2017: "According to the Ease of Language Understanding model (Rönnberg et al., 2013) working memory plays a critical role in perceiving speech stimuli in degraded listening conditions such as in background noise or in people with hearing loss (for an opposing view, see Füllgrabe and Rosen, 2016)."

LIP-READING, SPEECH INTELLIGIBILITY, AV-OLSA

Campbell, R. and Mohammed, T.J., **2010**. Speechreading for information gathering: a survey of scientific sources.
http://discovery.ucl.ac.uk/1347961/1/speechreading%20for%20information%20gathering.pdf
Great review on lipreading and factors. Verbal short term memory, familiarity with the talker, risk-taking personlity traits, age.
Full list of lipreading tests.

Fernandez-Lopez, Adriana, Oriol Martinez, and Federico M. Sukno. "Towards estimating the upper bound of visual-speech recognition: The visual lip-reading feasibility database." In **2017** 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 208-215. IEEE, 2017.
https://ieeexplore.ieee.org/abstract/document/7961743
"More importantly, it is not clear what would be the upper bound of visual-speech recognition, especially for systems not using context information (it has been argued that humans can read only around 30% of the information from the lips, and the rest is filled-in from the context [22], [32])"

Duchnowski, P., Lum, D.S., Krause, J.C., Sexton, M.G., Bratakos, M.S. and Braida, L.D., **2000**. Development of speechreading supplements based on automatic speech recognition. IEEE transactions on biomedical engineering, 47(4), pp.487-496.
https://ieeexplore.ieee.org/abstract/document/828148
Live sentences. CUNY and IEEE sentences.
"Live, rather than recorded, sentence productions were used"
"Strict scoring rules were used: responses considered correct had to agree in tense, number, and form with the sentence text. Homophonic responses were scored as correct"

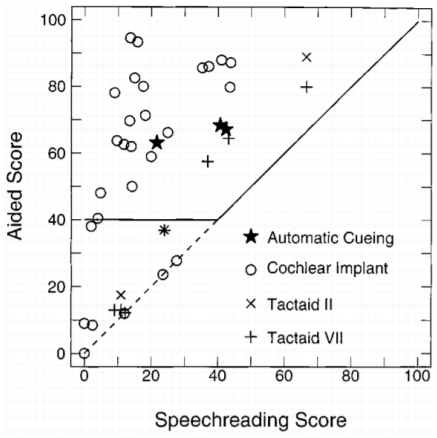


Fig. 2. Keyword scores with various speechreading aids versus unaided scores. Solid line delineates the region of significant benefit (speech reception allowing for reasonable conversation).

Lander, K. and Davies, R., **2008**. Does face familiarity influence speechreadability?. Quarterly Journal of Experimental Psychology, 61(7), pp.961-967.
https://www.tandfonline.com/doi/full/10.1080/17470210801908476
Yes, it does. 12 participants dropped out (out of 90) because the task was too difficult

"we have shown that a short period of familiarization time with the face and voice of the same speaker improves speechreading performance to a greater degree than does practice alone"
"The words were not presented silently but instead were presented at a very low level of volume, such that although the presence of the voice was audible (20 dB at 2 m), it was not loud enough to identify the presented words correctly. This was done to increase speechreading performance away from floor."

Baskent, D. and Bazo, D., **2011**. Audiovisual asynchrony detection and speech intelligibility in noise with moderate to severe sensorineural hearing impairment. Ear and hearing, 32(5), pp.582-592.
Audiovisual asynchrony

Woodhouse, L., Hickson, L. and Dodd, B., **2009**. Review of visual speech perception by hearing and hearing-impaired people: clinical implications. International Journal of Language & Communication Disorders, 44(3), pp.253-270.
https://www.tandfonline.com/doi/abs/10.1080/13682820802090281
Importance of including audiovisual speech tests in audiometry. Motivation for audiovisual speech tests.

CREATION AND EVALUATION OF AUDIOVISUAL MATERIAL

About normalizing the words. The parameters they use are the mean and SD of the SRT 50%? (dB SNR) and slope (%/dB). Usual values are -8dB SNR +-0.2 and 15%/dB +- 1%/dB. The probabilistic model (Kollmeier, 1990) makes predictions of post-normalization. They correct words with +-3 or 4 dB level adjustments. Word-specific SRT is also measured. Open and closed are also given (close-set usually lower SRT SNR).

Van Wanrooij, M.M., Van De Rijt, L.P., Roye, A., Mylanus, E.A. and Van Opstal, A.J., **2019**. The principle of inverse effectiveness in audiovisual speech perception. BioRxiv, p.585182.
https://www.biorxiv.org/content/biorxiv/early/2019/03/21/585182.full.pdf
https://www.frontiersin.org/articles/10.3389/fnhum.2019.00335/full
Dutch MST. Closed-set in quiet. Recorded the audiovisual material with 155 unique sentences. Tested at different SNR levels. They did not balance word intelligibility. No training (10 AV sentences). Average 50% intelligibility with visual-only. AO at -12, AV at -15 dB SNR (-3 dB SNR improvement).

Trounson, R.H., **2012**. Development of the UC Auditory-visual Matrix Sentence Test.
https://ir.canterbury.ac.nz/bitstream/handle/10092/10348/thesis_fulltext.pdf?sequence=1
Master thesis. Good read and summary of the whole topic. They mixed videos to create sentences. McClelland 2015 analized which ones were correctly mixed.

McClelland, A., **2015**. Refinement and Normalisation of the University of Canterbury Auditory-Visual Matrix Sentence Test.
https://ir.canterbury.ac.nz/bitstream/handle/10092/10456/thesis_fulltext.pdf?sequence=1&isAllowed=y
Two studies: 1) selection of mixed videos. 2) normalization of audio-only material.
“pair of compromise” (Brand & Kollmeier, 2002) → 20% and 80% (p1 = .19 and p2 = .81)
"adaptive procedure described by Brand and Kollmeier (2002)" → "an adaptive procedure: the preceding stimulus and listener’s response determine the presentation level of the following trial (Levitt, 1971)"
"Using a probabilistic model (Kollmeier, 1990)"

Jamaluddin, S.A., **2016**. Development and evaluation of the digit triplet and auditory-visual matrix sentence tests in Malay.
https://ir.canterbury.ac.nz/handle/10092/12288
PhD thesis.
4 lists 20 sentences AO. Lipreading without background noise. SNR adjusted to 30% SI from AO in the AV condition. The PHD is focused in AO, not in visual. It says that there are no floor effects in MMST-AV, when in VO the mean is 57.1% (page 266)???
"Participants were trained to a closed-set response method in auditory-only mode using 4 lists (80 sentences) and no training was given for the V and AV test modes"
"the signal-to-noise ratio level at which the participant would have responded at 30% of the words correctly." in audio-only.

Table 53: Auditory-visual and visual only scores for normal and hearing impaired listeners.

		Visual-only in quiet (%)	Auditory-visual in TSN (%)	Auditory-visual in BN (%)
Hearing impaired group	Mean	43.8	52.8	59.3
	σ	18.5	15.8	12.5
	Max	76	88.	100
	Min	16	25	38
Normal hearing group	Mean	57.1	66.0	70.4
	σ	17.9	12.9	13.5
	Max	84	86	94
	Min	25	45	44

"(1) the ability to lip read is independent of age and hearing levels; (2) hearing impairment affects auditory-visual speech perception and improving audition could improve speech perception; (3) the ability to integrate both auditory and visual cues is independent of age and hearing levels;"
"speech in noise is considered as the greatest handicap associated with hearing impairment (Smits, Kramer, et al., 2006; Walden, Schwartz, Williams, Holum-hardegen, & Crowley, 1983)."
"the use of a female speaker is encouraged as it is a compromise between the voice characteristics of a male and child speaker (Akeroyd et al., 2015)." "multi speaker analysis was done (Hochmuth, Jürgens, Brand, & Kollmeier, 2015)."
"Wardenga et al. (2015). noise level only for listeners with pure tone average of 47 dB HL. hearing levels exceeding 47 dB HL to be tested using the MST in quiet as no difference of testing with or without noise was observed in this group
"age has been found to have an effect on speech perception even when the hearing impairment is mild (Dubno et al., 1984) as well as in auditory-visual tasks (Tye-Murray et al., 2008, 2007c)."

O'Beirne, G.A., Trounson, R.H., McClelland, A.D., Jamaluddin, S. and Maclagan, M.A., **2015**. DEVELOPMENT OF AN AUDITORY-VISUAL MATRIX SENTENCE TEST IN NEW ZEALAND ENGLISH. Journal of International Advanced Otology, 11.
http://irep.iium.edu.my/66050/
Difficult to find the link, maybe results included in the PhD thesis.

Stone, J.M., **2016**. Evaluation of the University of Canterbury auditory-visual matrix sentence test.
https://core.ac.uk/download/pdf/35473612.pdf
Interesting read. Nevertheless they did not evaluate the visual part.

"communication handicap faced in daily life cannot be determined solely from measures of the loss of hearing sensitivity and disruption to speech understanding in quiet (Carhart & Young, 1976)"
"It has been suggested that the use of both word and sentence stimuli presented in noise provide more powerful information regarding the deficits faced by the client in real world listening environments (Grunditz & Magnusson, 2013)"
Realistic expectations: "(Humes, 1999; Taylor, 2003)."
Quantifying perceptual benefit SNR because of HA is beneficial for client assessment: (MacPherson & Akeroyd, 2014)
" a steep slope is thought to signify the sensitivity of a measure (Ozimek et al., 2010)"
Types of noise (Killion et al., 2004): speech shaped better for research (more control); babble noise more realistic (glimpses, masking release). Masking release small or absent in HI.
"is thought that as the ability to hear the signal deteriorates, the reliance on visual cues significantly increases (Tye-Murray et al., 2007b)"
"Exclusion of items that do not adequately fit the word-specific intelligibility function is generally advised (Kollmeier et al., 2015)"
Open and close set responses: "the findings are somewhat ambiguous between studies"
Page 38

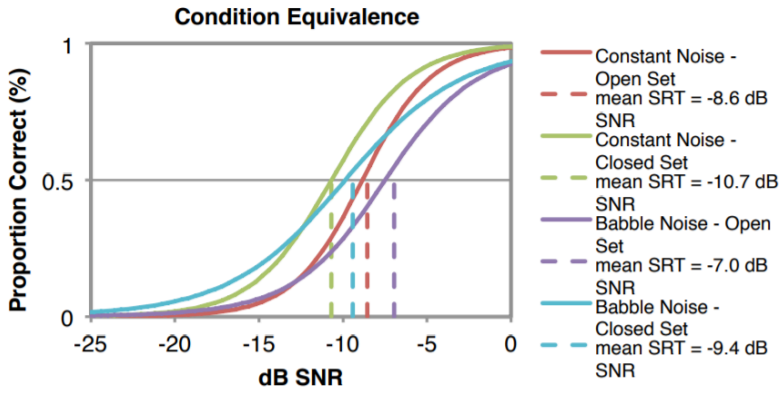


Figure 14. Intelligibility functions of each of the conditions of the UCAMST

Andre, E.R., **2016**. Is auditory-visual integration a factor in hearing aid outcomes?. https://ir.canterbury.ac.nz/handle/10092/12483
The relationship between AVE and hearing aid outcomes. 11 participants “moderate hearing impairment”, age average 77.9. Poor visual-alone recognition (VO-MST 24.5% mean, 8-44.7 range).

Variable	Units	Mean ± SD	Minimum	Maximum
SNR_20%	dB SNR	-4.2 ± 1.6	-6.8	-1.5
Auditory-visual score at SNR_20%	Percent	64.8 ± 11.4	41.3	82.7
Normalised AVE at SNR_20%	Percent	56.0 ± 14.2	26.7	78.0
SNR_80%	dB SNR	6.4 ± 4.0	1.0	14.7
Auditory-visual score at SNR_80%	Percent	90.4 ± 4.8	80.0	96.0
Normalised AVE at SNR_80%	Percent	52.0 ± 24.1	0.0	80.0
Visual-alone percent correct	Percent	24.5 ± 10.0	8.0	44.7

""“perceptual training” (Boothroyd, 2007). Such training can improve listening skills or improve the use of visual information along with improving listening (Katz et al., 2009). "
Older adults: "decreased ability to understand speech in background noise (Picou, Ricketts & Hornsby, 2013)."
Audiovisual integration: "combines an auditory signal and a visual signal to determine what words were said by a person speaking (Grant & Seitz, 1998)"
Most, Rothem, and Luntz (2009) auditory-visual, auditory-alone, and visual-alone. Showed AV higher scores than AO+VO. Audiovisual integration supported also by Tye-Murray, Sommers and Spehar (2007a). "
Integration provides benefits greater than simply the addition of the visual perception and auditory perception."
Tye-Murray (2007a) audiovisual speech perception: "1. The ability to lipread (understand speech using only visual information) 2. The ability to encode auditory information 3. The ability to integrate information obtained from the two modalities (auditory and visual)"
"Rogers (2012); Sommers, Tye-Murray and Spehar (2005); and Tye-Murray et al. (2008) also suggest that AVE tests using sentences provide results that differ from tests using consonants or individual words"
"Corthals et al. (1997) found that only factors 1 (SRT) and 2 (AVE) need be considered to predict real-world hearing disability"
" hearing impairment causes people to practise AVI more often and become better at AVI. This question has been researched with some conflicting findings."
"Tillberg et al. (1996) and Kyle et al. (2013) different findings with children-deaf-lipreading"
"The AVE score can be expressed as the difference between the percent correct score listening to speech-in-noise at the same SNR for the auditory-visual and auditory-alone conditions (Grant & Seitz, 1998). Alternatively, the AVE score can be expressed as the difference in SNR when listening to speech-in-noise, to achieve the same percent correct for the auditory-visual and auditory-alone conditions (MacLeod & Summerfield, 1987)"

Ripberger, A.R., **2018**. Further evaluation and validation of the University of Canterbury auditory-visual matrix sentence test.
https://ir.canterbury.ac.nz/handle/10092/15516
Only auditory

GENDER DIFFERENCES

Bench, J., Daly, N., Doyle, J. and Lind, C., **1995**. Choosing talkers for the BKB/A Speechreading Test: A procedure with observations on talker age and gender. British journal of audiology, 29(3), pp.172-187.
https://www.tandfonline.com/doi/abs/10.3109/03005369509086594
Selection of talkers/speakers. Younger female **judged** easier to lipread.

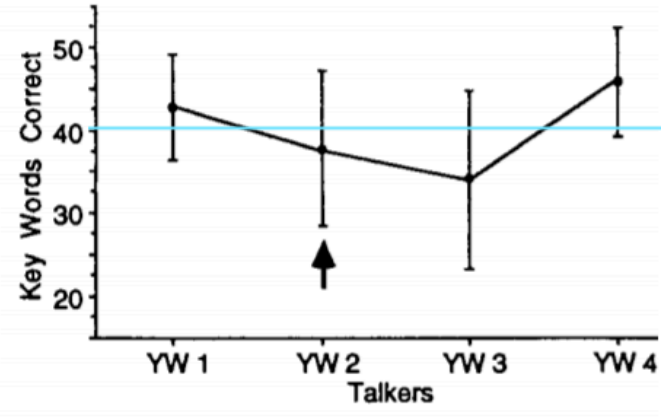


Fig. 4d. Mean key word scores for set YW at 60 d SPL noise level, with 1 SD bars (the arrow sho the selected talker)

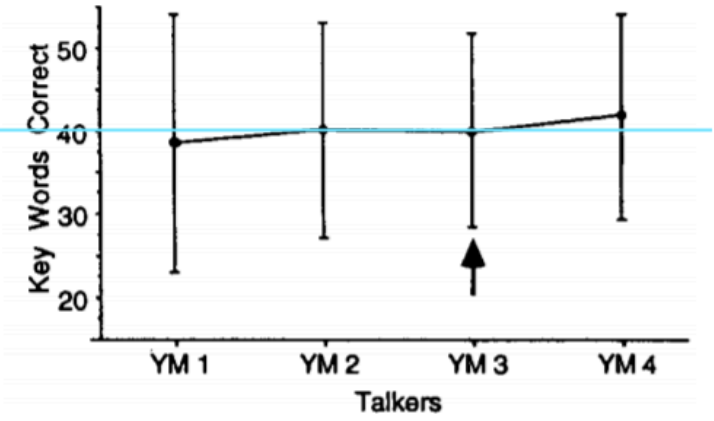


Fig. 4c. Mean key word scores for set YM at 60 dB SPL noise level, with 1 SD bars (the arrow shows the selected talker)

Daly, N., Bench, J. and Chappell, H., **1996**. Interpersonal impressions, gender stereotypes, and visual speech. Journal of Language and Social Psychology, 15(4), pp.468-479.
https://journals.sagepub.com/doi/abs/10.1177/0261927x960154004?casa_token=9fNTHxi3_jUAAAAA:RGSKALyI9pVPxIcHiO9bk-PCbyGBFy4DpWx6wOwBLVqI4oBByicBRzk5GdaeqYfmdP6nLVhycyaL
Need to read. Something about stereotypes? (Status, Solidarity and Performance)
"Female talkers have higher visual intelligibility than male talkers (Daly et al., 1996). This conflict can be explained by the fact that female talkers have to speak more clearly to be judged more educated or more competent, while male talkers are judged to be more competent and more educated than female talkers by virtue of being male."

Kricos, P.B., **1996**. Differences in visual intelligibility across talkers. In Speechreading by Humans and Machines (pp. 43-53). Springer, Berlin, Heidelberg.
https://link.springer.com/chapter/10.1007/978-3-662-13015-5_4
"The talker is a key variable affecting speechreading performance. Factors that appear to be related to speechreadability include a somewhat slower-than-normal rate of speech, precise articulation, appropriate gestures, and inclusion of appropriate pauses."

Kricos, P.B. and Lesner, S.A., **1982**. Differences in visual intelligibility across talkers. The Volta Review.
https://psycnet.apa.org/record/1984-14026-001
Same author as before. Only female talkers. "Results indicate that the number and nature of consonant visemes varied across talkers and were related to the talkers' sentence and intelligibility. It is suggested that teachers and clinicians evaluate and train lipreading skills using a variety of talkers, and that they objectively assess their own intelligibility through lipreading."

SI TESTS

Grunditz, M. and Magnusson, L., **2013**. Validation of a speech-in-noise test used for verification of hearing aid fitting. Hearing, Balance and Communication, 11(2), pp.64-71.
https://www.tandfonline.com/doi/full/10.3109/21695717.2013.782135
Stone 2016: "It has been suggested that the use of both word and sentence stimuli presented in noise provide more powerful information regarding the deficits faced by the client in real world listening environments (Grunditz & Magnusson, 2013)"

WFoster, J.R., Summerfield, A.Q., Marshall, D.H., Palmer, L., Ball, V. and Rosen, S., **1993**. Lip-reading the BKB sentence lists: Corrections for list and practice effects. British Journal of Audiology, 27(4), pp.233-246.
BKB is a speech intelligibility test in the UK (2nd most used in Clinics according to Stefan Wetzel). It has an audiovisual version.

van Laarhoven, T., Keetels, M., Schakel, L. and Vroomen, J., **2018**. Audio-visual speech in noise perception in dyslexia. Developmental science, 21(1), p.e12504.
https://www.frontiersin.org/articles/10.3389/fnhum.2019.00335/full
Word Recognition Error. "video recordings of 120 simple mono- and disyllabic Dutch nouns articulated by a female native Dutch speaker"

Xie, L., Wang, L. and Yang, S., **2016**. Visual Speech Animation. Handbook of Human Motion, pp.1-30.
https://www.researchgate.net/publication/312401529_Visual_Speech_Animation
Good introduction to the topic and the possible SNR improvements of hearing impaired.
"Sumby and Pollack found that seeing the speaker's face is equivalent to about 15dB signal-to-noise ratio (SNR) improvement of acoustic signal (Sumby and Pollack1954)"
"(Ostermann and Weissenfeld 2004) have shown that trust and attention of humans toward machines increase by 30% when communicating with a talking face instead of text only"

Devesse, A., Dudek, A., van Wieringen, A. and Wouters, J., **2018**. Speech intelligibility of virtual humans. International journal of audiology, 57(12), pp.914-922.
https://www.tandfonline.com/doi/pdf/10.1080/14992027.2018.1511922
Used videos and virtual characters to compute phoneme recognition (3 lists of 14 monosyllabic CVC words, 4 lists of 10 disyllabic words and 6 lists of 10 LIST sentences). They recorded their own video material (videos not found).
"The presence of the virtual speaker improved the speech intelligibility of sentences in noise by **1.5 to 2 dB**."
"male and female LIST (Jansen et al. 2014; van Wieringen and Wouters 2008)."
"LIST-sentences are comparable to the American English HINT- sentences (Nilsson, Soli and Sullivan 1994)"
"The sentences do not include questions, exclamations or proverbs and exist of two to three keywords (i.e. nouns, adjectives, numerals, adverbs and main verbs) and six to seven other words (i.e. articles, pronouns, prepositions, conjunctions, interjections and linking and auxiliary verbs)."
"Annosoft Lipsync Tool (Annosoft Lipsync Tool 2015, Version 4.0260),"
Video experiment: speech material: MS (monosyllabic) words, DS (disyllabic) words or sentences (LIST). Report all the phonemes. For the sentences, only the phonemes in the keywords were analyzed.

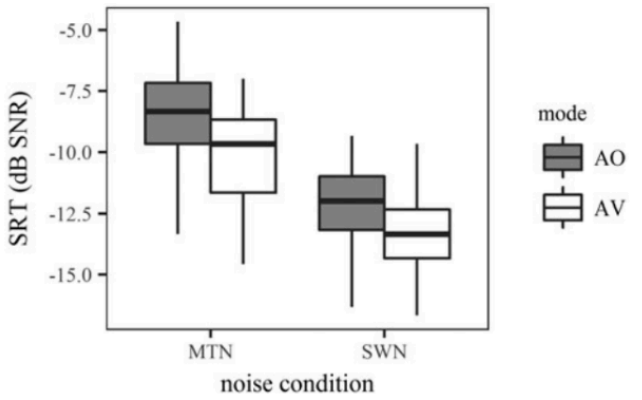


Figure 6. Boxplots representing the SRT's of the speech intelligibility experiment in different noise conditions (MTN: multitalker babble noise, SWN: speech weighted noise) and different listening modes (AO: Auditory-Only, AV: Auditory-Visual) for young normal hearing participants, $N = 35$.

AV = virtual characters, not video

"measured changes in SRT's, yielding benefits in the range of **2 to 6 dB** when video images are presented (Bernstein and Grant 2009; Brungart, Sheffield and Kubli 2014)."

Schreitmüller S, Frenken M, Bentz L, Ortmann M, Walger M, Meister H. Validating a Method to Assess Lipreading, Audiovisual Gain, and Integration During Speech Reception With Cochlear-Implanted and Normal-Hearing Subjects Using a Talking Head. Ear Hear. **2017** Oct
https://journals.lww.com/ear-hearing/Abstract/2018/05000/Validating_a_Method_to_Assess_Lipreading,.12.aspx
https://insights.ovid.com/crossref?an=00003446-201805000-00012
OLSA male with MASSY. There is a lot of related literature.
"visual-only Speech Reception with OLSA (**CI, 38%; NH, 12%**)"
"younger individuals obtained higher visual-only scores than older persons" → when lipreading
Training: "A total of 8 lists containing 20 sentences each were presented, therefore, comprising 4 lists in quiet (2 A-only and 2 V-only), as well as 4 lists in noise at 50% speech reception threshold level (2 A-only and 2 AV lists)." ... "speechreading performance is known to increase with practice and familiarity with a speaker (Lander & Davies 2008) "
"Desai et al. (2008), in contrast, is the only study we know to date that assessed VE using synthesized speech articulation."
"... This seems to indicate how at low auditory SR, the measure VE (visual enhancement) may strongly be triggered by lipreading abilities". Is then VE not happening at SRT 80%? They claim that no VE differences were found between CI and NH at TIL80%
"Both Bernstein et al. (2004) and Tye-Murray et al. (2011) showed that adding very simple visual structures (e.g., a Lissajous figure) [...] enhanced A-only SR."
"this face model in **three-fourth view**, which was the camera setting used for this study following recommendations for speechreading in multimedia settings (Campbell & Mohammed 2010)."
They use terms such as integration enhancement, visual enhancement and audio enhancement.

Desai, S., Stickney, G. and Zeng, F.G., **2008**. Auditory-visual speech perception in normal-hearing and cochlear-implant listeners. The Journal of the Acoustical Society of America, 123(1), pp.428-440.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2662523/

McGurk effect. "animated face (“Baldi”), which corresponded to the /ba/, /da/ and /ga/ sounds, were created using the Center for Spoken Language Understanding (CSLU) Speech Toolkit (Barnard et al., 2000; Massaro et al., 2000)."
Meister2017: "Desai et al. (2008), in contrast, is the only study we know to date that assessed VE using synthesized speech articulation. The authors found VE for phonemes articulated by Talking Head Baldi both for a group of CI recipients and NH listeners who were tested with vocoded speech."

Walden, B.E., Grant, K.W. and Cord, M.T., **2001**. Effects of amplification and speechreading on consonant recognition by persons with impaired hearing. Ear and hearing, 22(4), pp.333-341.
https://oce.ovid.com/article/00003446-200108000-00007/HTML
Are these results invalid, because word recognition does not reflect sentence recognition? (Andre 2016)
Andre 2016: "investigated the benefits of AVI and amplification"
"Consonant recognition" (VCV)
Andre2016: "AVI helped speech understanding most in the mid to high frequencies (place of articulation of speech, mostly for consonants) and amplification helped most in the low to mid frequencies (place and manner of articulation and voicing, mostly for vowels)"

Grant, K.W., Walden, B.E. and Seitz, P.F., **1998**. Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. The Journal of the Acoustical Society of America, 103(5), pp.2677-2690.
https://www.researchgate.net/profile/Ken_Grant/publication/13680577_Auditory-visual_speech_recognition_by_hearing-impaired_subjects_Consonant_recognition_sentence_recognition_and_auditory-visual_integration/links/5742078f08ae298602ee2867.pdf
Andre 2016: " In a within group army audiology study of 29 adults aged between 41 and 88 years (average age 65) with acquired SNHI, Grant and Seitz (1998) showed that, in theory, AVI training could improve auditory-visual consonant recognition by 26%. "

MacLeod, A. and Summerfield, Q., **1987**. Quantifying the contribution of vision to speech perception in noise. British journal of audiology, 21(2), pp.131-141.
https://www.tandfonline.com/doi/pdf/10.3109/03005368709077786
Andre 2016: "The range of speech reception threshold enhancement was from 6 dB to 15 dB between people and from 3 dB to 22 dB between sentences (some sentences are easier for AVI). MacLeod and Summerfield (1987) also found that AVE is highly correlated to lipreading (visual-alone) ability"

Corthals, P., Vinck, B., Vel, E.D. and Cauwenberge, P.V., **1997**. Audiovisual speech reception in noise and self-perceived hearing disability in sensorineural hearing loss. Audiology, 36(1), pp.46-56.
https://www.tandfonline.com/doi/pdf/10.3109/00206099709071960
"The conclusion is that auditory-only, audiogram-based prediction schemes must not be abandoned, but can be tailored to individuals on the basis of realistic audiovisual speech reception performance."
Plomp and Mimpen's sentences video tape recorded. Speaker's head and shoulders in a neutral background.

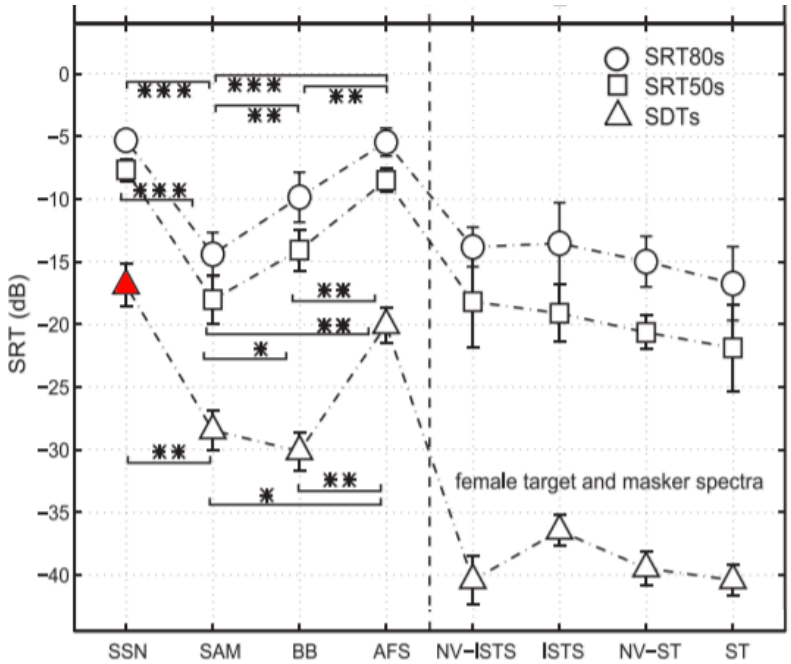
AUTOMATIC LIP-READING

Fernandez-Lopez, A. and Sukno, F.M., **2018**. Survey on automatic lip-reading in the era of deep learning. Image and Vision Computing, 78, pp.53-72.
https://www.sciencedirect.com/science/article/pii/S0262885618301276
Great survey. There are available AV databases and results and methods of machine learning algorithms for lip-reading.

Gorman, B.M. and Flatla, D.R., **2017**, May. A framework for speechreading acquisition tools. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 519-530).
https://www.researchgate.net/publication/325567216_A_Framework_for_Speechreading_Acquisition_Tools
Some tools for teaching lipreading

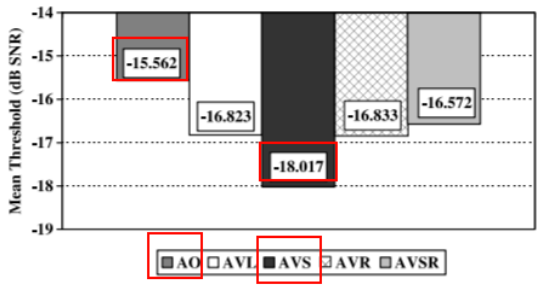
SPEECH DETECTION THRESHOLDS

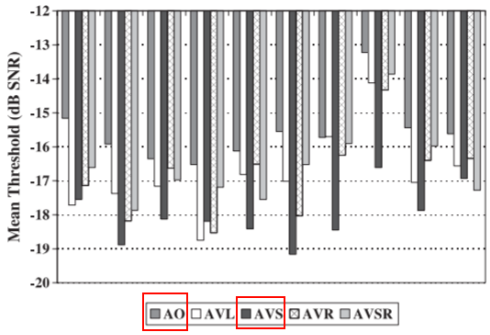
Schubotz, W., Brand, T., Kollmeier, B. and Ewert, S.D., **2016**. Monaural speech intelligibility and detection in maskers with varying amounts of spectro-temporal speech features. The Journal of the Acoustical Society of America, 140(1), pp.524-540.
https://asa.scitation.org/doi/full/10.1121/1.4955079
Speech detection threshold for female OLSA with SSN is -16.9 dB SNR.
"For the SSN and SAM-SSN maskers, SDTs were about 10 dB lower." than SRT50



Bernstein, L.E., Auer Jr, E.T. and Takayanagi, S., **2004**. Auditory speech detection in noise enhanced by lipreading. Speech Communication, 44(1-4), pp.5-18.

https://www.sciencedirect.com/science/article/abs/pii/S0167639304001165
Speech (/ba/) in white noise. -15 dB SNR for AO. -18.5 dB SNR for AV
"The mean improvement (AV AO threshold) in the detection threshold was 1.6 dB SNR (0.8–2.2 dB SNR) (Grant and Seitz, 2000)"





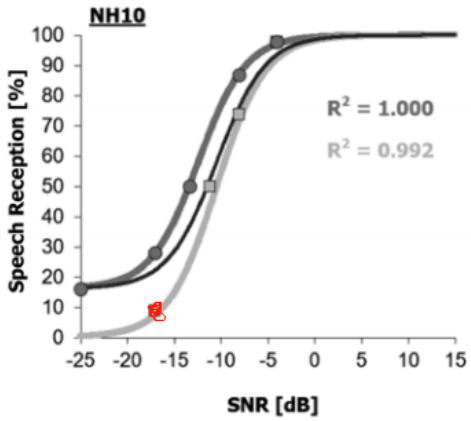
American Speech-Language-Hearing Association, **1988**. Determining threshold level for speech.
https://www.asha.org/policy/GL1988-00008.htm
"The speech detection threshold (in dB) should be consistent with the best pure tone threshold (in dB) between 250 and 4000 Hz (Olsen & Matkin, 1979) and should also be obtained at levels **8–9 dB weaker than the speech recognition threshold** (Chaiklin, 1959)." → Not specified if in quiet or in noise.

CONCLUSIONS

SRT 50% in OLSA male = -6.8 dB SNR
SRT 50% in OLSA female = -7.5 dB SNR (Schubotz 2016) or -9.4 dB SNR (Ahrlich 2013 - Bachelor Thesis).
Training effects!!

SDT —> should be 8-9 dB weaker than SRT50 (Chaiklin 1959)
SDT —> -17 dB SNR (min -18 dB SNR?) (Schubotz 2016)

At ~-17 dB SNR, 10% SRT for OLSA male in Schreitmüller 2017. Does not make a lot of sense, because SDT is at ~-14 dB SNR in Schubotz 2016.



EVALUATION OF LIP-SYNC

Crec que hi ha varies maneres d'evaluar el lipsync:
- Física: comparar diferències físiques entre la cara real i la cara virtual. [2]
- Subjectiva: evaluació subjectiva sobre el lip-sync [4]
- Comparativa: comparar amb altres sistemes [3]
- Speech intelligibility: evaluar si millora la intelligibilitat respecte audio-only input. [1]

[1] J. Liu, M. You, C. Chen, M. Song, “Real-time speech-driven animation of expressive talking faces,” International Journal of General Systems 40, pp. 439-455, 2011.
[2] L. Wei and Z. Deng, "A Practical Model for Live Speech-Driven Lip-Sync," in IEEE Computer Graphics and Applications 35(2), pp. 70-78, Mar.-Apr. 2015.
[3] Y. Xu, A. W. Feng, S. Marsella and A. Shapiro, ”A Practical and Configurable Lip Sync Method for Games,” In Proceedings of Motion on Games (MIG '13), pp. 131-140, 2013.
[4] S.L. Taylor, “Dynamic Units of Visual Speech,” in Proc. 11th ACM SIGGRAPH/Eurographics Conf. Computer Animation, pp. 275–284, 2012.

Personalment elegiria speech intelligibility, crec que és la més objectiva. Nosaltres ho provarem amb el sistema de lipsync que vam desenvolupar. Espero que fos això el que demanaves.

Summerfield, Q. (**1992**). Lipreading and audio-visual speech perception. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 335(1273), 71-78.
http://rstb.royalsocietypublishing.org/content/335/1273/71
Questions: What distinguishes better from poorer lipreaders? What are the effects of introducing a delay between the acoustical and the optical speech signals? What have attempts to produce computer animations of talking faces contributed to our understanding of the visual cues that distinguish consonants and vowels? How should the process of audio-visual integration in speech perce.ption be described; that is, how are the sights and sounds of talking faces represented at their conflux?

Lipsync: what is the minimal detectable asynchrony? What is the minimal asynchrony that disrupts the audio-visual intelligibility of speech?
McGrath Summerfield 1983 - minimal detectable asynchrony: 80ms (delay on lips) or 140ms (delay on audio).
Pandey et al. 1986 found 80 ms audio delay had no significant effect on SI.
Asymmetry? -> response latency in visual 20-100ms, in audio 1-4ms. → Sounds at 50m appear to be in sync, although audio is delayed due to travelling. Not for all subjects!

McGrath 1985, Summerfield 1989. Synthetic faces (very old - vector graphics): distinctiveness of vowels: 11 monophthongal british-english vowels in b-vowel-b syllables (beeb, bib, beb). 50-56% correct, with real video 78% correct.

Something about discordant gender audio-visual speech (female face, male voice) and the brain.

In addition, an increase in the size of the (mouth) opening raises the frequencies of each of the first three formants (Stevens & House 1955).

Massaro (1987) modelled subjects' identification responses to incompatible audio-visual combinations of consonants. Such responses were predicted more accurately by a model which preserved continuous values

of auditory and visual features up to the point of integration, than by a model which made categorical decisions about the phoneme presented in each modality.

We do not know what distinguishes better from poorer lipreaders. Intelligence and linguistic ability can be ruled out. Speed of low-level visual-neural processing correlates highly in some studies.

The 80 ms delay in audio is a motivation for HI processing.

Integration occurs before speech is categorized phonetically, but it is not known how the auditory and visual streams are represented at their conflux.

Linda Kozma-Spytek, Paula Tucker, and Christian Vogler. **2013**. Audio-visual speech understanding in simulated telephony applications by individuals with hearing loss. In Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '13)
https://dl.acm.org/citation.cfm?doid=2513383.2517032
Teleconference: audio ahead by 100ms bothers, audio delayed 100ms works the same as perfect audiovisual synchrony..

Ross K Maddox Huriye Atilgan Jennifer K Bizley Adrian KC Lee. **2015**. Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners.
https://elifesciences.org/articles/04995#abstract
Changing ball size with amplitude, frequency and other audio features. Received by Elif. "These findings suggest that in everyday conditions, where listeners can often see the source of a sound, temporal cues provided by vision can help listeners to select one sound source from a mixture." Might relate this to the lip-syncing.

Christian Phillipp Vogler
https://dl.acm.org/author_page.cfm?id=81100072347&coll=DL&dl=ACM&trk=0&cfid=840109976&cftoken=82427666
Some research with hearing impaired and telephony

Korin Richmond Lipsync
http://homepages.inf.ed.ac.uk/korin/sitenew/Publications/Inversion/index.html

Sascha Fagel and Caroline Clemens. 2003. Two Articulation Models for Audiovisual Speech Synthesis - Description and Determination
https://www.researchgate.net/publication/228912722_Two_articulation_models_for_audiovisual_speech_synthesis-description_and_determination
http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.58.6641
MASSY head used by Hartmut Meister

Schreitmüller S, Frenken M, Bentz L, Ortmann M, Walger M, Meister H. Validating a Method to Assess Lipreading, Audiovisual Gain, and Integration During Speech Reception With Cochlear-Implanted and Normal-Hearing Subjects Using a Talking Head. Ear Hear. **2017** Oct
http://europepmc.org/abstract/med/29068860
OLSA with MASSY

TALK FROM COLOGNE HARTMUT MEISTER

It would be interesting to compare different benefits from different avatars? Like percentages of improvement and make comparisons with real video and different avatars.

Some av vs aonly effects
priming - visual speech somewhat earlier/faster than auditory speech
Temporal coherence - a and v speech share similar envelope characteristics. Lexical segmentation and oscillatory entrainment
Articulatory

Speech audiometry
Monosyllables freiburger einsilber
Digits freiburger zahlen
Matrix-sentences (Oldenburger satztest)
Everyday sentences (Göttinger Satztest, HSM-Satztest)

Lipsync - articulation model for audiovisual speech synthesis fagel clemens 2004. Few blendshapes.

CI will have different benefits (fine structure vs envelope CIs)

Rationale

Grant & Walden 1998

(very!) schematic:

VOICING						fine-structure cues
Voiced					Unvoiced	
b,d,g,m,n,v,ð,z,ʒ,dʒ					p,t,k,f,θ,s,ʃ,tʃ	
MANNER OF ARTICULATION						envelope cues
Stop		Nasal	Fricative		Affricate	
b,p,g,k,d,t		m,n	v,f,ð,θ,z,s,ʒ,ʃ		dʒ,tʃ	
PLACE OF ARTICULATION						visual cues
Bilabial	Lingua-Velar	Lingua-Alveolar	Lingua-Dental	Lingua-Palatal	Labio-Dental	
b,p,m	g,k	d,t,n,s,z	ð,θ	ʒ,ʃ,dʒ,tʃ	v,f	

Eric Vatikiotis-Bateson, Takaaki Kuratate. 2012. Overview of audiovisual speech processing. In: Acoustical Science and Technology 33(3):135-141 · January 2012
https://www.researchgate.net/publication/275475518_Overview_of_audiovisual_speech_processing
"Visual information is found to contribute the most between these extremes, where some auditory information is available, but not too much."
"Other studies have found that intelligibility is enhanced so long as the visual information occurs at roughly the right time, modulating in time to the opening and closing of the vocal tract."

LIPSYNC INTERESTING

Speech graphics edinburgh
https://www.youtube.com/watch?v=wYwk07QM4rc

Speech Production and Articulation kNowledge group (SPAN) - USC
https://sail.usc.edu/span/resources.html
MRI databases of speech production. Really great.

Fabre, D., Hueber, T., Girin, L., Alameda-Pineda, X. and Badin, P., **2017**. Automatic animation of an articulatory tongue model from ultrasound images of the vocal tract. Speech Communication, 93, pp.63-75.
http://xavirema.eu/wp-content/papercite-data/pdf/Fabre-SpeechCom-2017.pdf
Database of images of the vocal track. Could be useful to generate lip-syncing

Higashimoto, T. and Sawada, H., **2002**. Speech production by a mechanical model: construction of a vocal tract and its control by neural network. In Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292) (Vol. 4, pp. 3858-3863). IEEE.
Utterances robot from Japan.
https://www.youtube.com/watch?v=HmSYnOvEueo&list=FLTZTDmnQcsqtWziKCZV6pyQ

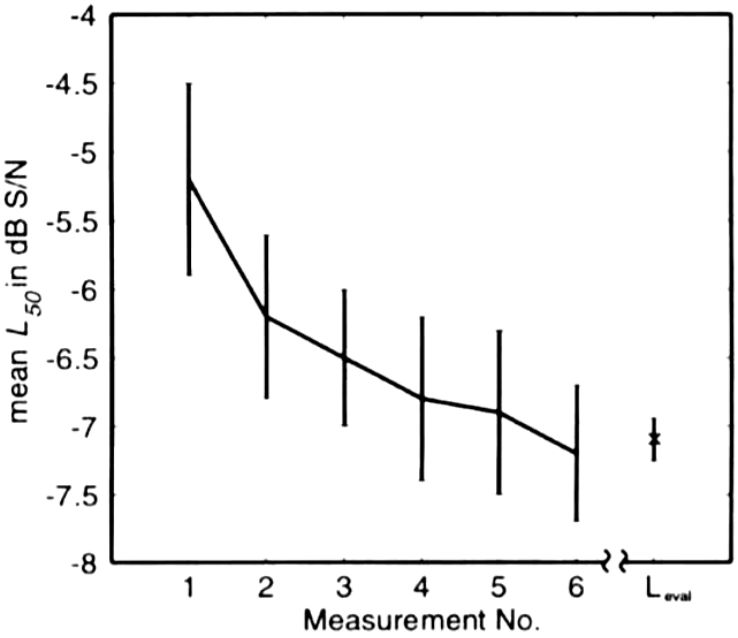
Oh, T.H., Dekel, T., Kim, C., Mosseri, I., Freeman, W.T., Rubinstein, M. and Matusik, W., **2019**. Speech2Face: Learning the Face Behind a Voice. arXiv preprint arXiv:1905.09773.
https://80.lv/articles/ai-reconstructs-faces-using-audio-recordings/?fbclid=IwAR3sMZlJf6dGlMGybTGPg3B3bGhi9KLmCbdmRZiXtQoqQtNu6HCTkblJ2uo
"AI Reconstructs Faces Using Audio Recordings"

OpenFace - Facial Action Units (FAPS, FACS)
https://github.com/TadasBaltrusaitis/OpenFace/

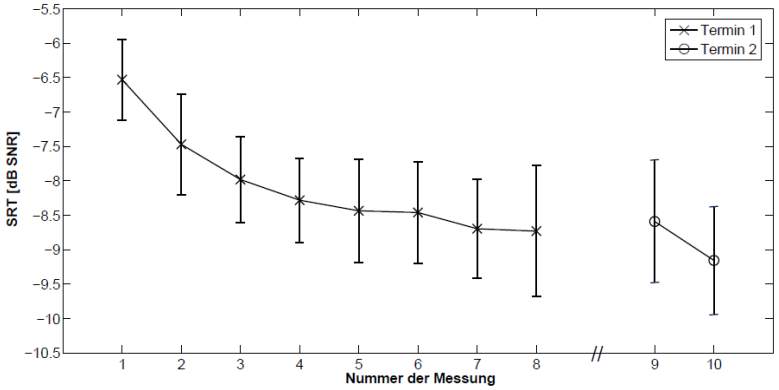
van Son, N., Huiskamp, T.M., Bosman, A.J. and Smoorenburg, G.F., **1994**. Viseme classifications of Dutch consonants and vowels. The Journal of the Acoustical Society of America, 96(3), pp.1341-1355.
https://asa.scitation.org/doi/10.1121/1.411324
"Vowel features in lipreading are lip rounding, degree of lip opening, and vowel duration, yielding the following visemes:" "lip rounding may roughly be related to the second formant, lip opening to the first formant."
"Lipreading expertise is not related to viseme recognition"
"consonant lipreading are lip articulation, degree of oral cavity opening, and place of articulation,"

OLSA

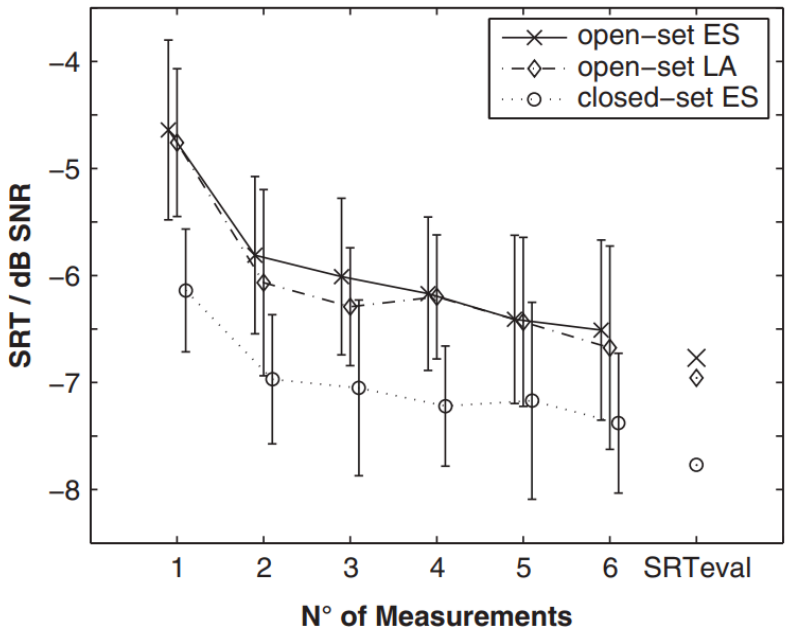
Wagener, K., Brand, T. and Kollmeier, B. **1999**. Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil III : Evaluation des Oldenburger Satztests
https://www.researchgate.net/publication/266735660_Entwicklung_und_Evaluation_eines_Satztests_fur_die_deutsche_Sprache_I_Design_des_Oldenburger_Satztests
https://www.uzh.ch/orl/dga-ev/ZfA_1999_38-2_044-056_Original.pdf
https://www.uzh.ch/orl/dga-ev/publikationen/zfaudiologie/archiv/ZfA_1999_38-3_086-095_Original.pdf



Ahrlich, M., **2013**. Optimierung und Evaluation des Oldenburger Satztests mit weiblicher Sprecherin und Untersuchung des Effekts des Sprechers auf die Sprachverständlichkeit. Bachelor Thesis
Provided by M. Zokoll. Female OLSA. From -6.5 dB SNR to -9.4 dB SNR. Open format. No citation?



Hochmuth, S., Brand, T., Zokoll, M.A., Castro, F.Z., Wardenga, N. and Kollmeier, B., **2012**. A Spanish matrix sentence test for assessing speech reception thresholds in noise. International Journal of Audiology, 51(7), pp.536-544.
https://www.tandfonline.com/doi/full/10.3109/14992027.2012.670731



Wagener K. **2004**. Factors Influencing Sentence Intelligibility in Noise, **2004**. (Dissertation). Oldenburg
http://oops.uni-oldenburg.de/460/1/wagfac04.pdf
Page 68 (general discussion)

Kollmeier B. **1990**. Messmethodik, Modellierung, und Verbesserung der Verst ä ndlichkeit von Sprache (in German). (Methodology, modeling, and improvement of speech intelligibility measurements). Habilitation thesis. Göttingen: University of Göttingen
Hochmuth et al. (2012): "The aim of the optimization procedure was to homogenize speech recognition across words by determining word-specific recognition curves and subsequently adjusting the word levels. This procedure generally increases the slope of the overall recognition curve (Kollmeier, 1990, see below)."

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M.A., Uslar, V., Brand, T. and Wagener, K.C., **2015**. The multilingual matrix test: Principles, applications, and comparison across languages: A review. International Journal of Audiology, 54(sup2), pp.3-16.
https://www.tandfonline.com/doi/full/10.3109/14992027.2015.1020971

Brand, T., & Kollmeier, B. (**2002**). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. The Journal of the Acoustical Society of America, 111(6), 2801-2810.

Nuesse, T., Wiercinski, B., Brand, T. and Holube, I., **2019**. Measuring Speech Recognition With a Matrix Test Using Synthetic Speech. Trends in hearing, 23, p.2331216519862982.
https://journals.sagepub.com/doi/full/10.1177/2331216519862982
Synthetic OLSA! They claim it can be used. I could not find out how they compute the word category scores, it might be good to ask.

NOISE vs QUIET

[Why did we do noise-quiet? Expected differences?]
Noise reasons found in [taylor2003]: hearing impaired complain in speech in noise situations, more realistic, "SNR loss"
Taylor 2003: "monosyllabic word lists presented in quiet are best relegated to the dustbin of history"
Stone 2016: "communication handicap faced in daily life cannot be determined solely from measures of the loss of hearing sensitivity and disruption to speech understanding in quiet (Carhart & Young, 1976)"
Jamaluddin2016: Killion & Niquette (2000) "the ability to hear in quiet is almost entirely independent of the ability to hear in noise"

Taylor, B., **2003**. Speech-in-noise tests: How and why to include them in your basic test battery. The Hearing Journal, 56(1), pp.40-42.
http://www.etymotic.com/media/publications/erl-0077-2003.pdf
"monosyllabic word lists presented in quiet are best relegated to the dustbin of history"

Hagerman, B. and Kinnefors, C., **1995**. Efficient adaptive methods for measuring speech reception threshold in quiet and in noise. Scandinavian audiology, 24(1), pp.71-77.
https://doi.org/10.3109/01050399509042213
"The ability of hearing-impaired patients to recognize speech may be divided into three parts, i.e. the speech reception threshold in quiet, in the following denoted SRT, the speech reception threshold in noise, in the following denoted S/N-threshold, and some speech recognition score at a comfortable speech level for monosyllabic words. The results of the first two (quiet and noise tests) are fairly independent of each other (Smoorenburg, 1992)"

Smoorenburg, G.F., **1992**. Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. The journal of the acoustical society of America, 91(1), pp.421-437.
https://asa.scitation.org/doi/abs/10.1121/1.402729
HI and NH SRT differences are bigger in quiet than in noise. Speech in quiet and in noise provide different information about the kind of hearing loss: sensorineural hearing loss (not distinguishing speech in noise), high-frequency/noise-induced hearing loss.
"The correlation between the speech thresholds in quiet and in noise was only R = 0.45"
"Speech perception in noise did not play an important part although there was general agreement that heating handicap should be evaluated in terms of ability to hear everyday speech under everyday conditions (Davis, 1965 [...] "It seems, however, that the decision not to include noisy environments was, at least partly, based upon an inability to decide upon a "representative" everyday noise" [...] "since they could not solve the question of a representative ambient noise, they concluded to consider everyday conditions to be listening in quiet"

Moradi, S., Lidestam, B. and Rönnberg, J., **2013**. Gated audiovisual speech identification in silence vs. noise: Effects on time and accuracy. Frontiers in Psychology, 4, p.359.
https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00359/full
Audiovisual speech recognition in noise and quiet. I don't understand the paradigm very well. IP = time required to identify a word. It takes longer to identify words/syllables in noise, but there is no effect in sentences.

"The findings are in line with our prediction that noise should not impact significantly on IPs or accuracy for final word identification in HP and LP sentences." [...] "it can be concluded that the provision of visual cues expedited IPs and increased the accuracy of speech stimuli identification in both silence and noise"
"Our results demonstrate that noise significantly delayed the IPs of audiovisually presented consonants and words. However, the IPs of final words in audiovisually presented sentences were not affected by noise, regardless of the sentence predictability level."
"In fact, the audiovisually presentation of speech stimuli **reduces working memory load** (i.e., Pichora-Fuller, 1996; Frtusova et al., 2013) which in turn eases processing of stimuli especially in noisy condition."

Behne, D.M., Wang, Y., Alm, M., Arntsen, I., Eg, R. and Valsø, A., **2007**, September. Changes in audio-visual speech perception during adulthood. In AVSP (p. 34).
https://pdfs.semanticscholar.org/6e0f/fdd1e78860911f3295759f5709c5d2a7b19a.pdf
I think they use the McGurk paradigm to see the effects of noise and quiet. In noise there is more audiovisual-fusion, thus they rely more on visual cues. "background noise (SNR=0dB) leads to more audiovisual fusion responses (e.g., [14, [15])."

SPEECH-SHAPED NOISE

[Why speech-shaped noise?] or test specific noise (TSN)
Stone2016: babble less reliable, speech-shaped noise more reliable (more precise results). (glimpses, masking release). Masking release small or absent in HI. Jamaluddin2016: "MacPherson & Akeroyd (2014) conducted a systematic review of 139 studies related to speech in noise tests. " "speech-type maskers resulted in shallower psychometric functions compared to amplitude noise masker or static maskers"
Jamaluddin2016: "Wagener & Brand (2005) better consistency when using speech-shaped noise""described no statistical difference between SRT for the two steady noises in subjects with varying levels of hearing.""found larger variations in SRT when fluctuating noise was used"
Stone 2016 found significant TSN-Babble noise differences.
More reliable, but not as realistic as babble noise. In babble noise there are glimpses and masking release, but not so present for HI. Thus babble noise, less reliable and comparable between NH and HI.

CLOSED-SET VS OPEN-SET TYPE

[Why did we do closed-open? Expected differences?]
Jamaluddin2016: "closed set responses are prone to training effects and guessing bias (Ozimek, Kutzner, & Libiszewski, 2012) whereas open set responses are not appropriate for listeners with disordered speech or limited language abilities". Closed-set absent experimenter.
Ozimek et al. (2010) no significant differences. Hochmuth et al. (2012) significant differences. But if taking into account training effects, results may not be valid. Stone 2016 found significant SRT differences.

TRAINING

Andre 2016: " with acquired SNHI, Grant and Seitz (1998) showed that, in theory, AVI training could improve auditory-visual consonant recognition by 26%. " "Lonka (1995) video vs real-person "The training improved test scores in this study from 28% to 38%"
Wagener2004: "The training effect was 2.2 dB and could be reduced to less than 1 dB, if two training lists of 20 sentences were performed prior to data collection"

Lidestam, B., Moradi, S., Pettersson, R. and Ricklefs, T., **2014**. Audiovisual training is better than auditory-only training for auditory-only speech-in-noise identification. The Journal of the Acoustical Society of America, 136(2), pp.EL142-EL147.
https://asa.scitation.org/doi/full/10.1121/1.4890200
Used hint sentences

INDIVIDUAL DIFFERENCES

AV benefit individual differences? Good lipreaders more benefit??

Wanrooij2019: "worse lipreaders benefited more from the additional auditory information for the audiovisually presented sentences". Said in other words, best lipreaders benefited less from auditory info in AV. But it does not say what we want: best lipreaders benefited more from visual information in AV.

GESTURES

Wagner, P., Malisz, Z., & Kopp, S. (**2014**). Gesture and speech in interaction: An overview. Speech Communication, 57, 209-232.
https://www.researchgate.net/publication/258696410_Gesture_and_speech_in_interaction_An_overview
http://www.sciencedirect.com/science/article/pii/S0167639313001295
Kendon (2004) distinguishes two main functions of co-speech gestures, namely substantial and pragmatic gestures. The former contribute to the utterance content, while the latter help negotiating aspects of the situational embedding.

Emblematic gestures, e.g. the “thumbs up” gesture, are conventionalized and language specific, while co-speech gesticulations are less standardized and work together with speech to accomplish communicative success.
Later, McNeill (2005) further refined the idea and argued for a complex of several continua, namely
(a) Continuum 1: relationship to speech (obligatory presence of speech – ... – obligatory absence of speech)
(b) Continuum 2: relationship to linguistic properties (linguistic properties absent – ... – linguistic properties present)
(c) Continuum 3: relationship to conventions (not conventionalized – ... – fully conventionalized)
(d) Continuum 4: character of semiosis (global & synthetic – ... – segmented & analytic)
Gesticulations are placed on the left ends of these continua (co-speech, no linguistic properties themselves, not conventialized, global meaning).

Gestures with the hands
Divided like BML sync points by Kendon (1972, 1980).
Gut and Milde (2003) pointed out that a function-oriented gestural phase classification, such as the one by Kendon above, differs from form-oriented descriptions of gestural phases.

The majority of gestures are performed in central gestural space, a sphere-shaped area in front of the speaker’s upper part of the body, below the neck and between the shoulders and elbows. Gestures occurring outside this area are said to be produced in peripheral gestural space and are believed to capture the addressee’s visual attention.

Semantic classification (McNeill 1992):
it is often useful to characterize and classify them with respect to their semantic function (alone, or along with formal features), e.g. using the classification by McNeill (1992):
(a) emblematic gestures bear a conventionalized meaning (”thumbs up”);
(b) iconic gestures resemble a certain physical aspect of the conveyed information, e.g. they may convey the shape of a described object or the direction of a movement;
(c) metaphoric gestures are iconic gestures that resemble abstract content rather than concrete entities (McNeill, 1992; Cienki and Mu¨ller, 2008);
(d) deictic gestures point out locations in space, with space often being of a conceptual rather than concrete nature;
(e) beat gestures are simple and fast movements of the hands (also called batons (Ekman and Friesen, 1972)). Rather than directly conveying meaning, they refer to the process of speaking itself by synchronizing with prosodic events in speech. They have been found to contribute to the perceived prominence of temporally aligned speech, and can function in the sense of Audiovisual Prosody (cf. Sections 2.2 and 4.2).

It is important to note that this classification should not be understood as defining distinct categories. McNeill (2005) convincingly argued that a simple functional classification of gestures is usually misleading.

Gesturing with the head
Recent analyses of multimodal corpora report that head nods predominate in gesturing when a listener role is assigned to the participants (81.5% of head movements were nods in Włodarczak et al., 2012 in a German active listening corpus) or in free conversation (56% in a multimodal corpus of Japanese in Ishi et al., this issue).

Translations not included in BML: Linear displacements (translations) are less frequently used, however, protruding and retracting the head along the X axis may correlate with displays of surprise and attention. We also proposed the inclusion of displacement along the horizontal axis, based on observations of German corpus data, as reported in Kousidis et al. (2013).

Hadar et al. (1985) noted that floor grabbing cues were usually expressed with wide and linear head movements (e.g. high amplitude, single nods), while synchronization with pitch accented syllables in the interlocutors speech occurred in cases of narrow, linear head gestures, e.g. low amplitude single nods. Rosenfeld and Hancks (1980) suggested that smaller, single nods often function as typical backchannels. While, according to Bousmalis et al. (2012), large amplitude, repeated nods are characteristic of affirmative meanings.

(Hadar et al. (1983)) A well-known result of this report is the incessant head movement found during speaking turns in his data: 89.9% of the time. The frequency ranges for communicative head movement were found to lie between 0.2–7 Hz.3 Within this range, movement amplitude was inversely correlated with frequency, i.e. the faster the movement, the more limited its excursion in space, with the exception of postural shifts, occurring typically with a high amplitude and high frequency movement.
FURTHER READ

Freigang, F., Kopp, S.: This is whats important – using speech and gesture to create focus in multimodal utterance. In: Lecture Notes in Computer Science, pp. 96–109 (**2016**)
Leo: "In the past, the problem has often been restricted to planning the non-verbal behavior of an agent [6, 12]."

OTHER INTERESTING

AUGMENTED REALITY

Dhruv Jain, Leah Findlater, Jamie Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E. Froehlich. **2015**. Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)

https://dl.acm.org/citation.cfm?id=2702393

Augmented reality, sound awareness, etc…

Developed the idea of showing positions of the sounds via AR glasses.

AUGMENTED SOUND REALITY

Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Hiipakka, J. and Lorho, G., **2004**. Augmented reality audio for mobile and wearable appliances. Journal of the Audio Engineering Society, 52(6), pp.618-639.

"For example, in some cases listeners found it very difficult to determine which sound sources in an augmented reality audio representation are real and which are virtual."

MCGURK EFFECT

Kristin J. Van Engen, Zilong Xie, Bharath Chandrasekaran. 2017. Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect. Attention, Perception, & Psychophysics February **2017**, Volume 79, Issue 2, pp 396–403

https://link.springer.com/article/10.3758/s13414-016-1238-9

https://csd.utexas.edu/sites/default/files/VanEngenetal2016_APP.pdf

McGurk effect doesn't exist in sentences. "Our results do not show a relationship between listeners' McGurk susceptibility and their ability to use visual cues to understand spoken sentences in noise, suggesting that McGurk susceptibility may not be a valid measure of audiovisual integration in everyday speech processing."

Strand, J., Cooperman, A., Rowe, J. and Simenstad, A., **2014**. Individual differences in susceptibility to the McGurk effect: links with lipreading and detecting audiovisual incongruity. Journal of Speech, Language, and Hearing Research, 57(6), pp.2322-2331.

https://www.researchgate.net/publication/266627543_Individual_Differences_in_Susceptibility_to_the_McGurk_Effect_Links_With_Lipreading_and_Detecting_Audiovisual_Incongruity

Needs further reading. On a quick look, McGurk effect and lipreading abilities are correlated.

"The links between MGS and lipreading also suggest that future studies seeking to quantify integration ability should measure lipreading skill to ensure that measures of MGS are separate from measures of visual-only identification."

VISUALS COMPETING SPEAKERS

Karen S. Helfer and Richard L. Freyman. **2005**. The role of visual speech cues in reducing energetic and informational masking. The Journal of the Acoustical Society of America 117, 842

https://doi.org/10.1121/1.1836832

Lip sync helps differentiate between competing speakers, better when the noise is speech noise and better when noise and target are in the same position.

"The data obtained here are consistent with the hypothesis that lipreading cues help to segregate a target voice from competing voices, in addition to the established benefit of supplementing masked phonetic information."

EARDRUMS MOVE
Kurtis G. Gruters, David L. K. Murphy, Cole D. Jenson, David W. Smith, Christopher A. Shera and Jennifer M. Groh. The eardrums move when the eyes move: A multisensory effect on the mechanics of hearing. **2018**. Proceedings of the National Academy of Sciences Jan 2018, 201717948; DOI: 10.1073/pnas.1717948115

WHEN LOOKING BETTER HEARING
Ross Maddox in ICA 2019
https://scholar.google.com/citations?hl=en&user=ky37necAAAAJ&view_op=list_works&sortby=pubdate

SOUND FLASH ILLUSION
Daniel P Kumpik, Helen E Roberts, Andrew J King, Andrew J King, Jennifer K Bizley, Jennifer K Bizley. **2014**. Visual sensitivity is a stronger determinant of illusory processes than auditory cue parameters in the sound-induced flash illusion June 2014Journal of Vision 14(7)
http://jov.arvojournals.org/article.aspx?articleid=2194009

Winn, M.B., Kan, A. and Litovsky, R.Y., 2019. Temporal dynamics and uncertainty in binaural hearing revealed by anticipatory eye movements. The Journal of the Acoustical Society of America, 145(2), pp.676-691.
https://asa.scitation.org/doi/full/10.1121/1.5088591
Not really interesting, just noting it down. Using some kind of ball to show where the sound will appear.

ASSISTANTS HELP SPEECH DISORDERS
Pradhan, A., Mehta, K. and Findlater, L., **2018**, April. Accessibility Came by Accident: Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (p. 459). ACM.
https://dl.acm.org/citation.cfm?id=3174033
"Findings show that, although some accessibility challenges exist, users with a range of disabilities are using the Amazon Echo, including for unexpected cases such as **speech therapy** and support for caregivers."
Keywords: IPA voice driven assistant speech disorder speak clearly clear speech

SPEECH RATE neq INFORMATION RATE
Coupé, C., Oh, Y.M., Dediu, D. and Pellegrino, F., **2019**. Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. Science advances, 5(9), p.eaaw2594.
https://advances.sciencemag.org/content/5/9/eaaw2594?utm_source=TrendMD&utm_medium=cpc&utm_campaign=TrendMD_1
Individual differences: slowest 4.3 syllables/s and the fastest one reaching 9.1 syllables/s on average

UNCANNY VALLEY
Kätsyri, J., Förger, K., Mäkäräinen, M. and Takala, T., **2015**. A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. Frontiers in psychology, 6, p.390.
https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00390/full
"the present review findings suggest that the uncanny valley **exists only under specific conditions.**"
"In particular, positive uncanny valley findings have been reported for stimuli in which the realism levels of **artificial and humanlike features are inconsistent with each other** (e.g., human eyes on an artificial face) or in which atypical features (e.g., grossly enlarged eyes) are present on humanlike faces."

WEARABLES WITH HEAD GESTURES

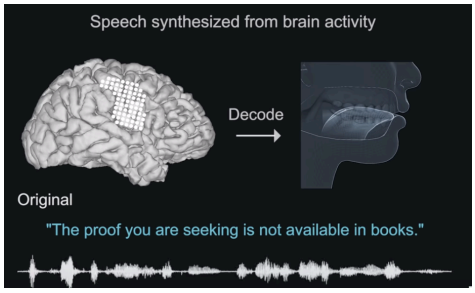
Hanheide, M., Bauckhage, C. and Sagerer, G., **2005**, October. Combining environmental cues & head gestures to interact with wearable devices. In Proceedings of the 7th international conference on Multimodal interfaces (pp. 25-31). ACM.
https://aiweb.techfak.uni-bielefeld.de/files/papers/Hanheide2005-CEC.pdf
This paper presents the prototype of a wearable, head-mounted device for advanced human-machine interaction that integrates speech recognition and computer vision with head gesture analysis based on inertial sensor data.

DATASETS VISUALIZATION
Matejka, J. and Fitzmaurice, G., **2017**, May. Same stats, different graphs: Generating datasets with varied appearance and identical statistics through simulated annealing. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 1290-1294). ACM.
Mean, SD and correlation are the same, but data looks very differently. **Datasaurus**.“dinosaurus”

SIGNIFICANCE OR SIGNIFICANTLY DIFFERENT
https://retractionwatch.com/2019/03/21/time-to-say-goodbye-to-statistically-significant-and-embrace-uncertainty-say-statisticians/?fbclid=IwAR2ZDsDEjoLSiDQUgZlDGcuhqpFPu2uY0OR18DJ3WX3fIYLLhY3_U6xbUkY
Wasserstein, R.L. and Lazar, N.A., **2016**. The ASA’s statement on p-values: context, process, and purpose. The American Statistician, 70(2), pp.129-133.
"...studies with multiple measurements on each participant or experimental unit. Such research presents the opportunity to p-hack your way to findings that likely have no scientific merit."

BRAIN SIGNALS TO RECONSTRUCT SPEECH
Akbari, H., Khalighinejad, B., Herrero, J.L., Mehta, A.D. and Mesgarani, N., **2019**. Towards reconstructing intelligible speech from the human auditory cortex. Scientific reports, 9(1), p.874.
https://gizmodo.com/neuroscientists-translate-brain-waves-into-recognizable-1832155006
Invasive electrocorticography (ECoG). Using machine learning. Based on speech.

Anumanchipalli, G.K., Chartier, J. and Chang, E.F., **2019**. Speech synthesis from neural decoding of spoken sentences. Nature, 568(7753), p.493.
https://www.nature.com/articles/s41586-019-1119-1
https://gizmodo.com/vocal-tract-simulator-translates-a-persons-brain-activi-1834268338
https://www.youtube.com/watch?v=kbX9FLJ6WKw
Intracranial electrodes. Using machine learning. Based on articulatory movements.



Jérémie Voix, In-Ear technologies, custom mold, protection, "The ear at the age of IoT"
https://youtu.be/8nykUOv2ZGw
https://youtu.be/npugr82KWXw
https://www.youtube.com/playlist?list=PLvTZNobdKFP31CcbsJ00T4SB64gsn90SJ

Seen at ICA2019 present by Jérémie Voix. Similar to Florian Denk's transparent ear piece. Facial recognition inside the ear, EEG around the ear, light therapy, menstruation control

STATISTICS
Automatic reporting of statistics with R (shared by Valerie Hazan)
https://easystats.github.io/report/index.html

AUDIOVISUAL INTEGRATION - PERCEPTION

Andre 2016: Audiovisual integration: "combines an auditory signal and a visual signal to determine what words were said by a person speaking (Grant & Seitz, 1998)"
Tye-Murray (2007a) audiovisual speech perception: "1. The ability to lipread (understand speech using only visual information) 2. The ability to encode auditory information 3. The ability to integrate information obtained from the two modalities (auditory and visual)
Most, Rothem, and Luntz (2009)

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (**2009**). The natural statistics of audiovisual speech. PLoS computational biology, 5(7), e1000436.
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000436
Very interesting paper that relates auditory and visual speech.

"The last observation from our measured statistics is that the timeto-voice (the delay between onset of mouth movements and the onset of the voice) is between 100 and 300 milliseconds."

"Therefore, under naturalistic conditions, structures mediating audiovisual speech perception must be tolerant to delays between visual and auditory signals. This observation is consistent with behavioral studies of audiovisual speech. Auditory speech must lag behind matching visual speech by greater than 250 milliseconds before any asynchrony is perceived by human subjects [68]. In stark contrast, detecting audio-visual asynchrony in artificial stimuli requires timing differences of as little as 20 milliseconds [69]."

Anja Gieseler, Maike Tahden, Maike TahdenChristiane Thiel, Christiane Thiel, Hans Colonius, Hans Colonius. **2018**. Does hearing aid use affect audiovisual integration in mild hearing impairment?
https://www.researchgate.net/publication/323233875_Does_hearing_aid_use_affect_audiovisual_integration_in_mild_hearing_impairment/references
Sound induced flash illusion. Oldenburg people. Quite extensive references.
"Timed-up and-go test (Podsiadlo and Richardson 1991), respectively. The latter measures the time participants need to stand up from a chair, walk a 3-m distance and back, and sit down again. When less than 10 s are required, mobility is considered unimpaired. "
"Visual acuity and functional mobility were evaluated with the Visus-C (Haase and Hohmann 1982) "

Jesse, A. and Janse, E., **2012**. Audiovisual benefit for recognition of speech presented with single-talker noise in older listeners. Language and cognitive processes, 27(7-8), pp.1167-1191.
Phonemes, maybe should add the reference because of individual differences
"Audiovisual benefits for older adults were similar in size to those observed for younger adults in terms of response latencies, but smaller for detection accuracy. Older adults with more hearing loss showed larger audiovisual benefits."

Mishra S. , Lunner T. , Stenfelt S. , Rönnberg J. & Rudner M . **2013** . Seeing the talker's face supports executive processing of speech in steady state noise . Front Syst Neurosci , 7 : 96 . DOI: 10.3389/fnsys.2013.00096 .
McGarrigle et al.: Such interference, however, may be modulated by the presence of a visual aid (i.e. seeing the talker's face), which has been found to reinforce the executive processing of speech in steady state noise (Mishra et al, 2013).

Rudner, M., Mishra, S., Stenfelt, S., Lunner, T., & Rönnberg, J. (**2016**). Seeing the Talker's Face Improves Free Recall of Speech for Young Adults With Normal Hearing but Not Older Adults With Hearing Loss. Journal of Speech, Language, and Hearing Research, 59(3), 590-599.

Ciraolo, M.F., O'Hanlon, S.M., Robinson, C.W. and Sinnett, S., **2020**. Stimulus onset modulates auditory and visual dominance. Vision, 4(1), p.14.
https://www.mdpi.com/2411-5150/4/1/14/htm
Overall, these results provide empirical support for the hypothesis that the modality that engages processing first is more likely to show dominance, and suggest that more explicit measures of sensory dominance may favor the visual modality.

Schutte, M., Ewert, S.D. and Wiegrebe, L., **2019**. The percept of reverberation is not affected by visual room impression in virtual environments. The Journal of the Acoustical Society of America, 145(3), pp.EL229-EL235.
https://asa.scitation.org/doi/full/10.1121/1.5093642
Rift DK2, RAZR, Blender.
Non Congruent visual and audio. Interesting literature.

HMD PROBLEMS (head-mounted displays)

Vergence Accommodation Mismatch. Light-fields, monovision blur
https://youtu.be/c8Ge08MwSLQ?t=8m29s

Hoffman, D.M., Girshick, A.R., Akeley, K. and Banks, M.S., **2008**. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. Journal of vision, 8(3), pp.33-33.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2879326/

Assymetry IPD
https://www.youtube.com/watch?v=addUnJpjjv4

Niehorster, D.C., Li, L. and Lappe, M., **2017**. The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research. i-Perception, 8(3), p.2041669517708205.
http://journals.sagepub.com/doi/abs/10.1177/2041669517708205
Can be discussed that SteamVR is constantly improving:
https://steamcommunity.com/app/250820/allnews/#scrollTop=8748

EEG - attention

Bleichner, M.G., Mirkovic, B. and Debener, S., **2016**. Identifying auditory attention with ear-EEG: cEEGrid versus high-density cap-EEG comparison. Journal of neural engineering, 13(6), p.066004.
http://iopscience.iop.org/article/10.1088/1741-2560/13/6/066004/meta
"sound streams were three seconds long and differed in the direction of origin (front, left, right) and the number of beats (3, 4, 5 respectively), as well as the timbre and pitch"
"the direction of attention could be decoded significantly above chance (50%) for at least 16 out of 20 participants for both systems" (EEG and cEEGrid)

Mirkovic, B., Bleichner, M.G., De Vos, M. and Debener, S., **2016**. Target speaker detection with concealed EEG around the ear. Frontiers in neuroscience, 10, p.349.
https://www.frontiersin.org/articles/10.3389/fnins.2016.00349/full
cEEGrid "with 69.3% accuracy, while cap-EEG signals resulted in the accuracy of 84.8%."
"lesser-known fairy tales narrated in German language" 10 min duration. They processed the signals so the lower competing voice would be raised, as if to simulate Lombard speech (only amplitude).
Nogueira 2019: "Only two of their 18 NH participants did not reach accuracies above chance level, whereas in the current study only 5 out of 10 NH"

Nogueira, W., Dolhopiatenko, H., Schierholz, I., Buechner, A., Mirkovic, B., Bleichner, M.G. and Debener, S., **2019**. Decoding Selective Attention in Normal Hearing Listeners and Bilateral Cochlear Implant Users With Concealed Ear EEG. Frontiers in neuroscience, 13, p.720.
https://www.researchgate.net/publication/334554089_Decoding_Selective_Attention_in_Normal_Hearing_Listeners_and_Bilateral_Cochlear_Implant_Users_With_Concealed_Ear_EEG
About EEG and CI, also included cEEGrid. Only working for 50% of the participants, pretty discouraging.
"Similar to previous studies (Mirkovic et al., 2015; Nogueira et al., 2019) two German narrations (‘A drama in the air’ by Jules Verne narrated by a male speaker and ‘Two brothers’ by the Grimm brothers, narrated by a female speaker"
"The large difference between the cEEGrid performance and the scalp EEG performance using a decreased number of electrodes can be explained by the position of the reference electrode. (PDF) Decoding Selective Attention in Normal Hearing Listeners and Bilateral Cochlear Implant Users With Concealed Ear EEG. " [...]
"the nose tip is used as reference, which is located far away from the recording electrodes and therefore gives higher amplitudes."

Pomper, U. and Chait, M., **2017**. The impact of visual gaze direction on auditory object tracking. Scientific reports, 7(1), p.4640.
https://www.ncbi.nlm.nih.gov/pubmed/28680049
"Subjective experience suggests that we are able to direct our auditory attention independent of our visual gaze, e.g when shadowing a nearby conversation at a cocktail party. But what are the consequences at the behavioural and neural level? While numerous studies have investigated both auditory attention and visual gaze independently, little is known about their interaction during selective listening."

Ed Lalor's work

ATTENTION

Macdonald, J.S. and Lavie, N., **2011**. Visual perceptual load induces inattentional deafness. Attention, Perception, & Psychophysics, 73(6), pp.1780-1789.
https://link.springer.com/content/pdf/10.3758/s13414-011-0144-4.pdf
“inattentional deafness”.

Virginia Best, Erol J. Ozmeral, Barbara G. Shinn-Cunningham. **2007**. Visually-guided Attention Enhances Target Identification in a Complex Auditory Scene. Journal for the Association for Research in Otolaryngology June 2007, Volume 8, Issue 2, pp 294–304
http://www.cns.bu.edu/~shinn/resources/pdfs/2007/2007JARO_Best.pdf
Bird sounds stimuli. Normal hearing. These results suggest that in real world situations, information about where a target of interest is located can enhance its identification, while information about when to listen can also be helpful when targets are unfamiliar or extremely similar to their competitors.

LOMBARD SPEECH SYNTHESIS

López, A.R., Seshadri, S., Juvela, L., Räsänen, O., Alku, P. (**2017**) Speaking Style Conversion from Normal to Lombard Speech Using a Glottal Vocoder and Bayesian GMMs. Proc. Interspeech 2017, 1363-1367, DOI: 10.21437/Interspeech.2017-400.
http://isca-speech.org/archive/Interspeech_2017/abstracts/0400.html
Normal to Lombard conversion. Naturalness was evaluated. Open source code??

Michael Fitzpatrick, Jeesun Kim, Chris Davis. **2013**. Auditory and Auditory-Visual Lombard Speech Perception by Younger and Older Adults
http://avsp2013.loria.fr/proceedings/papers/paper_37.pdf
Stimuli is vowel-consonant-vowel (VCV) or CVC. Lombard AV is better than Normal AV for younger adults, but not for older adults

Eric Vatikiotis-Bateson, Victor Chung, Kevin Lutz, Nicole Mirante, Jolien Otten, and Johanna Tan. **2006**. Auditory, but perhaps not visual, processing of Lombard speech
http://asa.scitation.org/doi/10.1121/1.4786950
Contradictory findings from Fitzpatrick. Couldn't access the article! Interesting to check the stimuli presented

Chris Davis, Jeesun Kim, Katja Grauwinkel , Hansjörg Mixdorff. **2006**. Lombard speech: Auditory (A), Visual (V) and AV effects
http://isle.illinois.edu/sprosig/sp2006/contents/papers/PS3-24_0252.pdf
Might be interesting for modelling head movement.
"Lombard speech showed increased AV speech correlations between RMS speech intensity and the PCs that involved jaw and mouth movement. A similar increased correlation occurred for intensity and head rotation (pitch). For Lombard speech, all talkers showed an increased correlation between F0 and head translation (raising and lowering). Increased F0 correlations for other head movements were more idiosyncratic."

Simon Alexanderson, Jonas Beskow. **2014**. Animated Lombard speech: Motion capture, facial animation and visual intelligibility of speech produced in adverse conditions
Audiovisual lombard speech! Facial retargeting

REALISTIC NOISE LEVELS ENVIRONMENTAL

Weisser, A. and Buchholz, J.M., **2019**. Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions. The Journal of the Acoustical Society of America, 145(1), pp.349-360.
https://asa.scitation.org/doi/pdf/10.1121/1.5087567?class=pdf
"conversations were elicited between pairs of people by asking them to play a collaborative game"; "with acoustic background noise at above approximately 69 dB SPL at 1 m distance, or 75 dB SPL at 0.5 m, the average SNR can become negative."

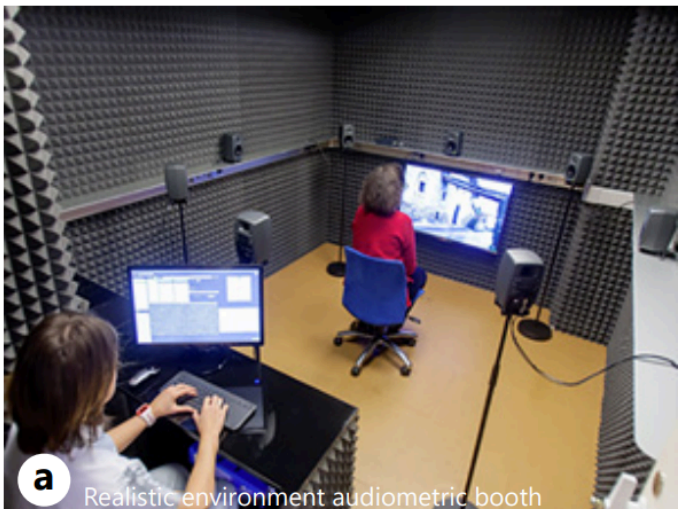
Gregory S. Farber, and Lily M Wang. **2017**. Analyses of crowd-sourced sound levels of restaurants and bars in New York City
https://asa.scitation.org/doi/pdf/10.1121/2.0000674
dbA levels for restaurants and cafeterias. 78 to 81 dbA

Smeds, K., Wolters, F., and Rung, M. (**2015**). " Estimation of signal-to-noise ratios in realistic sound scenarios," J. Am. Acad. Audiol. 26, 183–196.
https://doi.org/10.3766/jaaa.26.2.7
SRNs levels for hearing impaired in realistic situations, usually higher than +5dB. Results are discussed in John F. Culling 2016.

John F. Culling. **2016**. Speech intelligibility in virtual restaurants. The Journal of the Acoustical Society of America 140, 2418 (2016);
https://asa.scitation.org/doi/10.1121/1.4964401
Quite good summary of realistic simulation of environments.
"Simple cases (one interfering voice) permit SRTs of around as low as −10 dB, but more complex cases can elevate SRTs to −5 dB."
Smeds 2015: "The most striking result was that SNRs tended to be +5 dB or greater, suggesting that the frequent discussion of negative SNRs in the literature may be misguided. There are, however, a number of caveats that one should consider with respect to this finding."
"**No study to date, has attempted to include visual information in a realistic listening simulation"**
Yi et a 2013.: "Nonetheless, effects on performance of competition from "distracter" faces have been observed (Yi et al., 2013), suggesting that truly realistic results can only be obtained with audio-visually rendered interferers."

AUDIOVISUAL SIMULATIONS FOR HEARING RESEARCH - Similar to Gesture lab

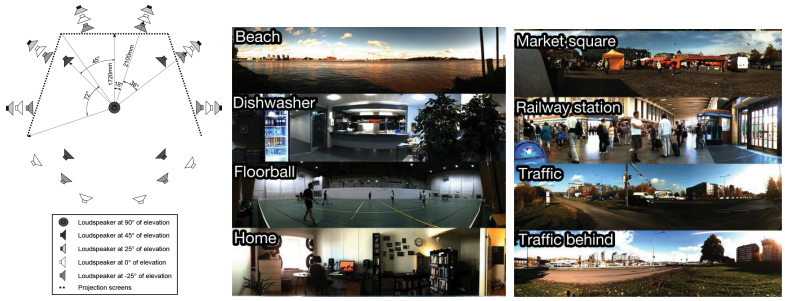
Vigliano, M., Huarte, A., Borro, D., Lasarte, U., & Manrique Rodriguez, M. J. (**2021**). Realistic Environment Audiometric Booth: Development and Clinical Validation. Audiology and Neurotology, 26(5), 317-326. https://doi.org/10.1159/000512759
https://navarra.okdiario.com/articulo/sociedad/audicion-pionero-proyecto-navarra/20191129143128300291.html
Vigliano, M., (**2021**) Cabina de Audiometría en Ambientes Reales
https://dadun.unav.edu/handle/10171/63265
20 scenarios, 8 loudspeakers, 3D videos. Different tests (speech in noise, discomfort levels, speech in quiet).



Seeber, B.U. and Clapp, S.W., **2017**. Interactive simulation and free-field auralization of acoustic space with the rtSOFE. The Journal of the Acoustical Society of America, 141(5), pp.3974-3974.
https://asa.scitation.org/doi/10.1121/1.4989063
Abstract only, people from Munich

Annelies Devesse, Alexander Dudek, Astrid van Wieringen & Jan Wouters. **2017.** The AVATAR-approach: How real-life listening affects speech intelligibility and listening effort
Financed by OTICON. I will hear more about her soon.
Realistic 3D environments over a big 2D flat projection. Several virtual characters and dual-task paradigm. Speech in noise.

Pulkki and Rummukainen
Pulkii is the creator of VBAP and DirAC. Rummukainen is doing research with audiovisual environments with a similar system to ours. They basically use 360º camera + DirAC.



Database Audiovisual environments
http://research.spa.aalto.fi/publications/papers/natural-scene-discrimination/

Olli Rummukainen and Catarina Mendonça. **2016**. Reproducing reality: Multimodal contributions in natural scene discrimination. ACM Trans. Appl. Percept. 14, 1, Article 1 (July 2016), 19 pages.
https://dl.acm.org/citation.cfm?id=2915917
https://www.researchgate.net/publication/305441630_Reproducing_Reality_Multimodal_Contributions_in_Natural_Scene_Discrimination
Measures subjective responses for movement, openness and noisiness in audio only, visual only and audiovisual.

Rummukainen, O., Radun, J., Virtanen, T. and Pulkki, V., **2014**. Categorization of natural dynamic audiovisual scenes. PloS one, 9(5), p.e95848.
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0095848
About the environments (beach, bells, cafeteria, classroom, cyclicsts, dishwasher, floorbal, hall, home, lecture hall, market square, narrow view, quiet street, railway station, subway station, talkers in street, traffic, traffic behind, tram)

Bolaños, J.G. and Pulkki, V., **2012**, April. Immersive audiovisual environment with 3D audio playback. In Audio Engineering Society Convention 132. Audio Engineering Society.
http://www.aes.org/e-lib/browse.cfm?elib=16242
Paper about the laboratory

Pulkki, V., Politis, A., Laitinen, M.V., Vilkamo, J. and Ahonen, J., **2017**. First-order directional audio coding (DirAC). Parametric Time-Frequency Domain Spatial Audio, pp.89-138.
https://www.researchgate.net/publication/320400479_First-Order_Directional_Audio_Coding_DirAC
"The system performs well if the recorded spatial sound matches the implicit assumptions about the sound field in DirAC, that at each frequency band only a single source is dominant at one time, with a moderate level of reverberation. In cases where the recorded sound field strongly violates these assumptions, audible distortions, artifacts, may occur. The artifacts were pronounced with a high number of loudspeakers, and especially in listening conditions with low reverberation."

Vilkamo, J., Lokki, T. and Pulkki, V., **2009**. Directional audio coding: Virtual microphone-based synthesis and subjective evaluation. Journal of the Audio Engineering Society, 57(9), pp.709-724.
http://www.aes.org/e-lib/browse.cfm?elib=14838
Subjective evaluation DirAC. "Subjective evaluations in both an anechoic chamber and a listening room are also presented. In a direct comparison with a reference scenario generated by using a large number of loudspeakers, the average perceived quality of reproduction of DirAC was excellent or good in almost all listening conditions. DirAC fared well, without problems, for all loudspeaker configurations tested."

Rummukainen, O., **2016**. Reproducing reality: Perception and quality in immersive audiovisual environments. https://aaltodoc.aalto.fi/handle/123456789/23541
Thesis. Experiments: openness, noisiness and movement with audio only, visual only, audiovisual; audio and video width affect the perceived degradation of a stimulus; others (read abstract of thesis)

"Bech [118], when using a standard television set. He found that increasing audio reproduction width increases the quality of reproduction of space."

Rummukainen, O. and Pulkki, V., **2012**, July. Audiovisual reproduction in surrounding display: Effect of spatial width of audio and video. In Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on (pp. 127-132). IEEE.
https://ieeexplore.ieee.org/abstract/document/6263861/
Spatial audio with and video with comparisons. "The results show both audio and video width affect the perceived degradation of a stimulus. The effect of audio width decreases as video width is decreased. Constrained correspondence analysis suggests the reasons for highest perceived degradation to be wrong audio direction, reduced video width and missing essential content."

Politis, A., McCormack, L. and Pulkki, V., **2017**, October. Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing. In Applications of Signal Processing to Audio and Acoustics (WASPAA), 2017 IEEE Workshop on (pp. 379-383). IEEE.
https://ieeexplore.ieee.org/abstract/document/8170059/
Improved DirAC (**for binaural reproduction?**). "Therefore, this paper proposes an improved DirAC method that directly synthesises the binaural cues based on the estimated spatial parameters. This method can accommodate higher-order Ambisonics (HOA) signals and has reduced computational requirements; thus, making it suitable for lightweight processing with fast update rates and head-tracking support. According to listening tests, when utilising only FOA signals, the method results in equivalent or higher spatial accuracy than third-order Ambisonics and significantly outperforms FOA reproduction."

Pulkki, V., **2007**. Spatial sound reproduction with directional audio coding. Journal of the Audio Engineering Society, 55(6), pp.503-516.

Gupta, R., Lam, B., Hong, J.Y., Ong, Z.T., Gan, W.S., Chong, S.H. and Feng, J., **2017**. 3D AUDIO AR/VR CAPTURE AND REPRODUCTION SETUP FOR AURALIZATION OF SOUNDSCAPES.
https://www.researchgate.net/publication/318851185_3D_AUDIO_ARVR_CAPTURE_AND_REPRODUCTION_SETUP_FOR_AURALIZATION_OF_SOUNDSCAPES
360 camera, ambisonics mic, no evaluation. Oculus Rift and HRTF binaural.

Hong, J.Y., He, J., Lam, B., Gupta, R. and Gan, W.S., **2017**. Spatial audio for soundscape design: Recording and reproduction. Applied Sciences, 7(6), p.627.
https://www.researchgate.net/publication/317617940_Spatial_Audio_for_Soundscape_Design_Recording_and_Reproduction
Full review of capture and reproduction techniques of audio. Doesn't speak about engines and virtual scenes. Ecological validity of ambisonics: "In a study for comparing the ecological validity among the different reproduction methods related to physical reconstruction of sound, Gustavino et al. [16] compared the stereophonic and ambisonic reproduction techniques. The study analyzed verbal data collected through questionnaires,and compared it to the field survey using semantic differential analyses for different sound

samples.The study concluded that a “neutral visual environment” along with spatial immersion of recreated soundstage is essential to ensure high ecological validity"
"a disjoint in the spatialness of the audio and visual elements will degrade the ecological validity [87]."

VIRTUAL SCENES ACOUSTIC REAL-TIME ENGINES

Poirier-Quinot, D., Katz, B. and Noisternig, M., **2017**, September. EVERTims: Open source framework for real-time auralization in architectural acoustics and virtual reality. In 20th International Conference on Digital Audio Effects (DAFx-17).
https://hal.archives-ouvertes.fr/hal-01712896/
Integrated acoustic engine in Blender Game Engine
"As the reflection paths can only be computed up to a limited order in real-time, a statistical model, based on Sabine's reverberation time formula, is used to approximate the late reverberation time of the current room. The reverberation time is estimated per octave band and sent to the auralization unit. A complete characterisation of the EVERTims modeller unit, including performance assessment and an evaluation of its simulation fidelity as compared to other auralization engines, is presented in [6]"

Remaggi, L., Kim, H., Jackson, P.J. and Hilton, A., **2018**, October. An Audio-Visual Method for Room Boundary Estimation and Material Recognition. In Proceedings of the 2018 Workshop on Audio-Visual Scene Understanding for Immersive Multimedia (pp. 3-9). ACM.
https://dl.acm.org/citation.cfm?id=3264876
Useful to reconstruct scenes for TASCAR (geometry model and material properties of the room). They use two 360 cameras, a omni-mic and do octave-band sweeps with a loudspeaker.

Savioja, L. and Svensson, U.P., **2015**. Overview of geometrical room acoustic modeling techniques. The Journal of the Acoustical Society of America, 138(2), pp.708-730.
https://asa.scitation.org/doi/full/10.1121/1.4926438
"A typical room impulse response, as measured in a room, or obtained by room acoustical simulation, is illustrated in Fig. 1(a). [...] Such responses can be considered to consist of three separate parts: direct sound, early reflections, and late reverberation."
"The main trend in the development of hybrid methods has been to utilize image-source or beam tracing techniques to detect the early specular reflection paths and various radiosity-based techniques for the late reverberation as there is no need for the accurate detection of the specular reflections to obtain the later component of the impulse response (see Sec. II A); instead, a computationally more efficient but less accurate technique can be used"

FIND DOI TOOL
Simple Text Query
https://doi.crossref.org/simpleTextQuery

PUPILLOMETRY

Kret, M.E. and Sjak-Shie, E.E., **2018**. Preprocessing pupil size data: Guidelines and code. Behavior research methods, pp.1-7.
https://link.springer.com/article/10.3758/s13428-018-1075-y?wt_mc=alerts.TOCjournals&utm_source=toc&utm_medium=email&utm_campaign=toc_13428_51_3

Hershman, R., Henik, A. and Cohen, N., **2019**. CHAP: Open-source software for processing and analyzing pupillometry data. Behavior research methods, pp.1-16.
https://link.springer.com/article/10.3758/s13428-018-01190-1?wt_mc=alerts.TOCjournals&utm_source=toc&utm_medium=email&utm_campaign=toc_13428_51_3

Geller, J., Winn, M., Mahr, T. and Mirman, D., **2019**. GazeR: A Package for Processing Gaze Position and Pupil Size Data.
https://psyarxiv.com/gvcxb

McGarrigle, R., Rakusen, L. and Mattys, S., **2020**. Effortful listening under the microscope: examining relations between pupillometric and subjective markers of effort and tiredness from listening. PsyArXiv. June, 12.
https://psyarxiv.com/kych5

ENGAGEMENT

Nadine Glas, Catherine Pelachaud. **2015**. Definitions of Engagement in Human-Agent Interaction. 2015 International Conference on Affective Computing and Intelligent Interaction (ACII)
https://www.researchgate.net/publication/282863531_Definitions_of_Engagement_in_Human-Agent_Interaction
Quite interesting review about engagement, its definitions and measurements. Talks about O'Brien, but many other papers about engagement when interacting.
"There is no consensus on the measure of engagement. Some see it as an instance, others as a value or degree, a process, a description, etcetera"
"Within the cases where engagement relates to a state we can distinct two subcategories: those that assign a **binary value** to engagement (presence or absence, e.g. instance Def. 1) and those that imply a **continuous notion of engagement** (e.g. value Def. 2, degree Def. 8)."

"we can state that engagement in human agent interaction is a **complex concept**"

O'Brien and MacLean. **2009**. Measuring the User Engagement Process. Engagement by Design PreConference Workshop, CHI 2009 Digital Life New World, Boston, MA, April 5, 2009
http://faculty.arts.ubc.ca/hobrien/files/OBrien_MacLean_2009_Measuring_the_User_Engagement_Process.pdf
Engagement may be viewed as both an outcome of experience and a process during an interaction [O'Brien MacLean 2009].

To date, the measurement of engagement has involved self-report measures, namely questionnaires [3; 10; 22]. Some researchers have used performance indicators, not as measures of engagement, but as correlates of engagement.

Measurement in HCI The measurement of human-computer interactions has typically centered on usability, defined as search effectiveness (i.e. ability to find the information sought, the user's judgment of the relevance or merit of search results), search efficiency (e.g. time to complete the search task), and user satisfaction (i.e. users' affective assessment of the task outcome).

Aesthetic Appeal, Novelty, Focused Attention, Felt Involvement, Perceived Usability, and Endurability (i.e. willingness to use an application 3 again or recommend it to others) [15; 16].

Performance indicators and physiological metrics have the advantage of recording users' physiological reactions or behaviors over the course of an interaction with a system. Yet while these metrics answer the question of what is taking place and the users' endogenous physiological states, they do not address users' expressed cognitive or emotional dispositions, both of which are critical to understanding engagement. Thus, usability metrics, biometrics, and eye tracking data will be more powerful when examined in concert with users' self reports and other observational data.

Heather L. O'Brien and Elaine G. Toms. **2010**. The development and evaluation of a survey to measure user engagement. Journal of the American Society for Information Science and Technology
http://onlinelibrary.wiley.com/doi/10.1002/asi.21229/epdf
Notably, Perceived Usability played a mediating role in the relationship between Endurability and Novelty, Aesthetics, Felt Involvement, and Focused Attention.

According to this body of work, engagement is impacted by: media richness through the use of animations and video (Webster & Ahuja, 2006); format (e.g., text, audio, and video) (Chapman, 1997; Jacques et al., 1995; Laarni, Ravaja, Kallinen, & Saari, 2004); interactivity and exploration (Haywood & Cairns, 2005); communication or socialization with others (Haywood & Cairns, 2005; Hull & Reid, 2003); aesthetics and sensory appeal (Haywood & Cairns, 2005; Hull & Reid, 2003; Laarni et al., 2004); intellectual challenge (Douglas & Hargadon, 2000; Mandryk, 2004; Said, 2004); and affective involvement (Laarni et al., 2004; Schraw et al., 1998; Said, 2004).

Performance and usability metrics could be employed with respect to engagement, but not in the traditional sense since the Perceived Usability factor of the scale incorporated **effectiveness** and **satisfaction**, but not **efficiency**. This makes sense since an engaged user may not be concerned about how much time they are spend-ing using an application. However, it does mean that some usability metrics, including time spent performing a task or examining a screen, may not be applicable.

Performance measures

Some researchers have used performance indicators, not as measures of engagement, but as correlates of engagement. For example, Konradt and Sultz (2001) employed pre- and post-task measures to examine changes in users' affective and cognitive states over the course of interacting with an educational application. Chapman (1997) studied the relationship between engagement (as measured by Webster and Ho's instrument) and performance metrics, such as time on task (as gathered through log files) and knowledge of system content (as measured by multiple-choice quizzes) as a consequence of media format. One researcher posited that multiple measures (self-report, participant observation, biometrics) are needed to study engagement but did not offer any results of empirical data collection or describe how these measures might be triangulated to say something meaningful about engagement (Champion, 2003).

Erik Frøkjær, Morten Hertzum, and Kasper Hornbæk. **2000.** Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '00)

http://dl.acm.org/citation.cfm?id=332455

Unless domain specific studies suggest otherwise, effectiveness, efficiency, and satisfaction should be considered independent aspect of usability and all be included in usability testing.

Florin Dobrian, Vyas Sekar, Asad Awan, Ion Stoica, Dilip Joseph, Aditya Ganjam, Jibin Zhan, and Hui Zhang. **2011.** Understanding the impact of video quality on user engagement. SIGCOMM Comput. Commun. Rev. 41, 4 (August 2011)

http://dl.acm.org/citation.cfm?id=2018478

Engagement Metrics: play time. every subject has a unique profile measured by number of views and total play time.

ENGAGEMENT MEASURES

Performance indicators, physiological data and self-report measures

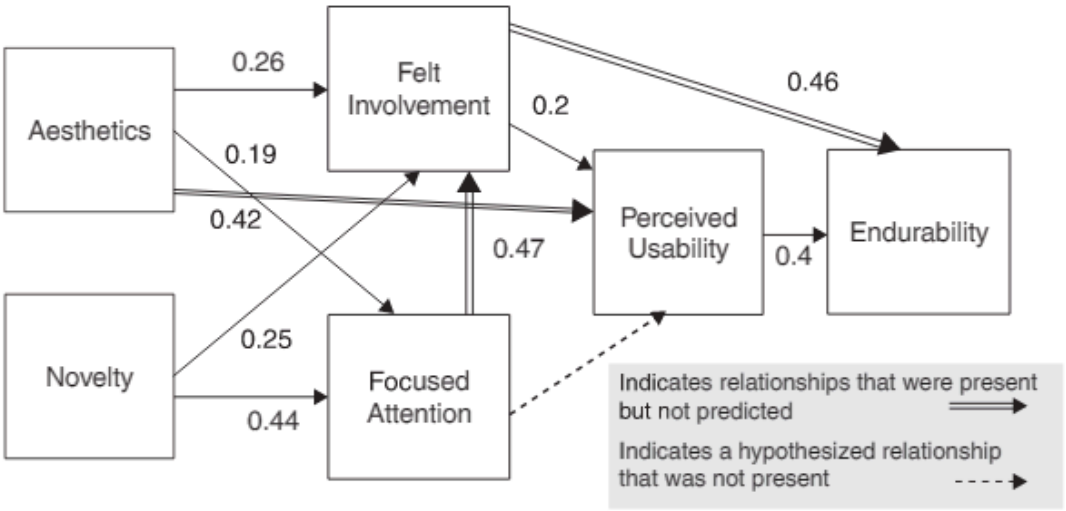
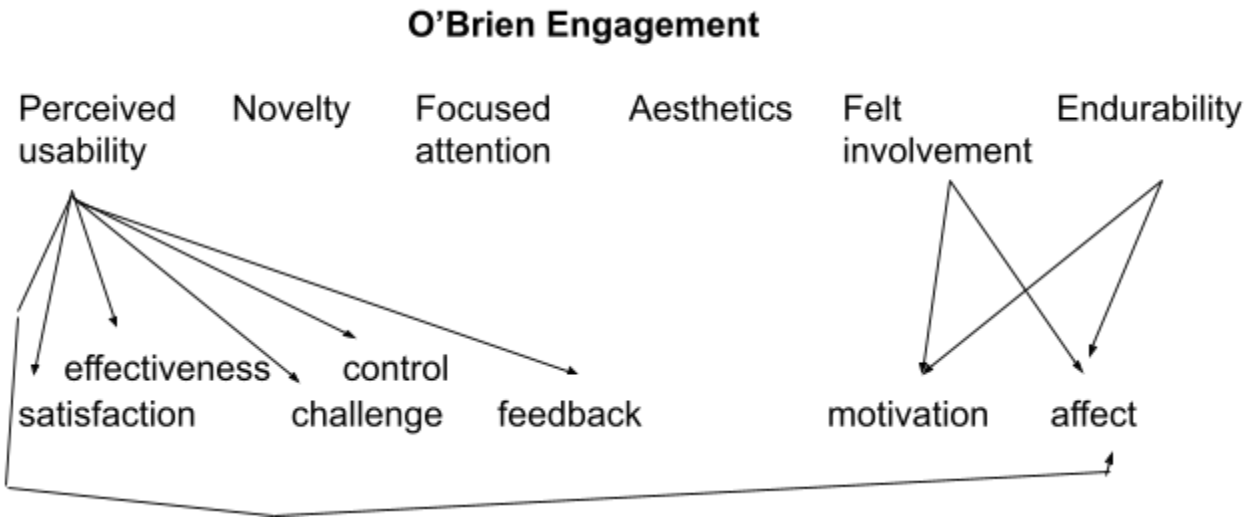


FIG. 3. Final structural equation model.



TAM - Technology Acceptance Model

Fred D. Davis. **1989**. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly

https://www.researchgate.net/profile/Fred_Davis2/publication/200085965_Perceived_Usefulness_Perceived_Ease_of_Use_and_User_Acceptance_of_Information_Technology/links/54ad66dc0cf24aca1c6f3765.pdf

Viswanath Venkatesh , Michael G. Morris , Gordon B. Davis , Fred D. Davis. **2003**. User acceptance of information technology: toward a unified view

http://nwresearch.wikispaces.com/file/view/Venkatesh+User+Acceptance+of+Information+Technology+2003.pdf

Judy van Bilijon, Karen Renaud. **2008**. A Qualitative Study of the Applicability of Technology Acceptance Models to Senior Mobile Phone Users

Questionnaires in Usability Engineering A List of Frequently Asked Questions (3rd Ed.)
http://www.ucc.ie/hfrg/resources/qfaq1.html
Useful information about questionnaires in general

LABORATORY VS REALITY COMPARISONS

Maempel, H.J. and Horn, M., **2018**. Audiovisual perception of real and virtual rooms. Journal of Virtual Reality and Broadcasting, 14(5).
https://www.researchgate.net/publication/325381535_Audiovisual_perception_of_real_and_virtual_rooms
https://www.jvrb.org/past-issues/14.2017/4668
Big questionnaire with many aspects.
160º cylindrical screen stereoscopic - binaural dynamic synthesis through extraaural headphones
"To date, no studies comparing VEs and REs were found that systematically investigated purely auditory features such as loudness and reverberance, and few comparative studies have investigated aesthetic features based on both optical and optoacoustic stimuli"
"Providing optoacoustic stimuli, Bishop & Rohrmann [BR03] asked participants to complete an extensive questionnaire covering cognitive and affective aspects. They found that ratings of pleasure, naturalness, appreciation, and general realism (among others) were significantly decreased in theVE; **familiarity**, however, was rated higher."
Terminology of **optoacustic** instead of audiovisual: "[...] it was designed strictly according to previously elaborated methodological criteria [Mae17]which are formulated by means of a consistent terminology (e.g.domain vs.modality,optical vs.visual, acoustic vs.auditory,property vs.feature,unimodal vs.intermodal vs.supramodal) used also in this article. Audiovisual perception of real and virtual rooms [...]"

Bishop, I.D. and Rohrmann, B., **2003**. Subjective responses to simulated and real environments: a comparison. Landscape and urban planning, 65(4), pp.261-277.
http://realities.id.tue.nl/wp-content/uploads/2010/03/bishop-rohrmann-2003.pdf
They recreate a walk in a street environment and make the people walk in the same environment. Evaluation is done through subjective questionnaires. The real street junction is this:
https://www.google.com/maps/search/Camberwell+junction+melbourne/@-37.832626,145.0586731,236a,35y,157.03h,37.65t/data=!3m1!1e3 They did a very good job in the recreation taking into account that it was done in 2002. " “Familiarity” is greater for the simulation than for the site visit. It may be that the ‘generic’ quality of the trees and other textures reduces the individuality of the place and it becomes more ‘familiar’ in a general sense. Thus, there seems to be a difference in the interpretation of the term ‘familiar’ in the two conditions: in

the real environment people say ‘not familiar’ if they have not been there before. Whereas, in the simulation, people may say ‘familiar’ if they have seen this type of place before."
Cited by Maempel and Horn 2018.

Rummukainen, O., **2016**. Reproducing reality: Perception and quality in immersive audiovisual environments. https://aaltodoc.aalto.fi/handle/123456789/23541
Section 2.7 Sensory integration and quality evaluation (page 32)
Chapter 4. Quality of Immersive Experiences (page 45)

Rummukainen, O., Wang, J., Li, Z., Robotham, T., Yan, Z., Li, Z., Xie, X., Nagel, F. and Habets, E.A., **2018**, October. Influence of visual content on the perceived audio quality in virtual reality. In Audio Engineering Society Convention 145. Audio Engineering Society.
https://www.researchgate.net/publication/328449132_Influence_of_visual_content_on_the_perceived_audio_quality_in_virtual_reality
Comparison of three audio systems (Oculus audio, Resonance and Unity) and three conditions of visual content (full content, half of the content, empty space). Evaluated audio quality with a slider (poor to excel·lent).
Section 2.7 Sensory integration and quality evaluation (page 32)
Chapter 4. Quality of Immersive Experiences (page 45)

Joseph, A., Browning, M. H., & Jiang, S. (**2020**). Using immersive virtual environments (IVEs) to conduct environmental design research: A primer and decision framework. HERD: Health Environments Research & Design Journal, 13(3), 11-25.
https://journals.sagepub.com/doi/abs/10.1177/1937586720924787
"The ecological validity of virtual experiences is commonly measured with the user’s reported level of presence (Cummings et al., 2012)." "It has been argued that the declaration of ecological validity is impossible (Dole & Ju, 2019)." Talks about ecological validity and formulates the problem. The examples are related previous publications about pre-op rooms, stress recovery and EEG validity. Important concepts to keep in mind: ecological validity, visual realism, behavioral realism, and contextual realism. It has nice images about the VR setups.
Page 5.

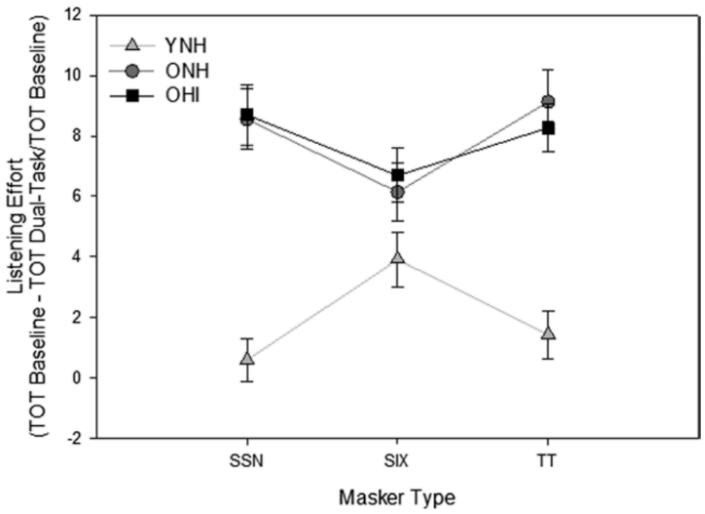
Dole, L., & Ju, W. (**2019**, May). Face and ecological validity in simulations: lessons from search-and-rescue HRI. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-8).
https://dl.acm.org/doi/pdf/10.1145/3290605.3300681
https://doi.org/10.1145/3290605.3300681
An interesting paper talking about ecological validity in emergency-rescue simulations. They support the presence questionnaires as a tool. There are some practical guidelines at the end.
"When dealing with regular simulations like the living room example, one should also estimate ecological validity by correlating laboratory and naturalistic data"
"Ideally, presence should be measured using a combination of participants’ self-report, behavior, and physiology."

DUAL TASK LISTENING EFFORT BALL

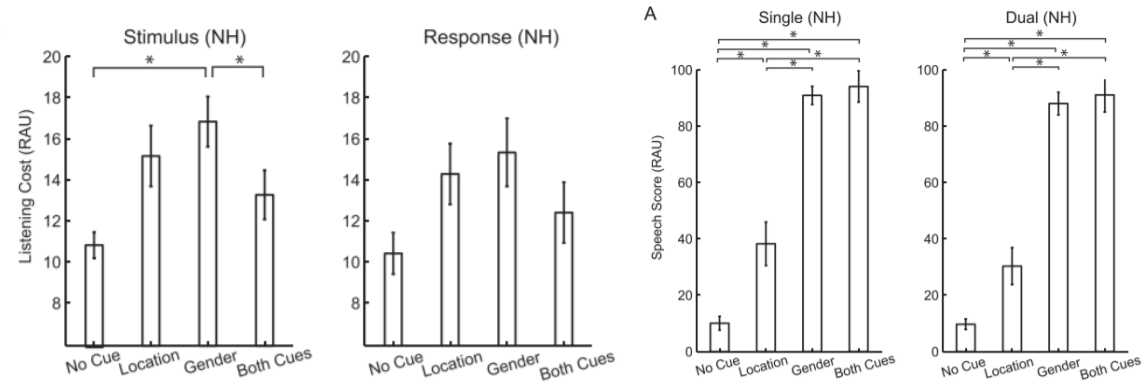
Gagne, J.P., Besser, J. and Lemke, U., **2017**. Behavioral assessment of listening effort using a dual-task paradigm: A review. Trends in hearing, 21, p.2331216516687287.
https://journals.sagepub.com/doi/pdf/10.1177/2331216516687287
Good summary of different dual-task experiments.
"It is difficult to determine whether the secondary task that consisted of the word-category recognition task was more sensitive to listening effort because it involved processing linguistic information (just as the primary task did) or because it was cognitively more demanding in other respects."
"Nevertheless, the results of the investigation do suggest that the type of secondary task used may have an influence on whether or not listening effort can be measured."

"At the present time, there is no clear indication concerning which of the two approaches is the most appropriate to characterize listening effort. However, it is noteworthy that Whelan (2008) claims that in some experiments, the **RT** data collected may not meet some of the assumptions required to use an analysis of variance to test for significant effects." "Both **accuracy** and **RTs** were assessed. As a first approximation, both methods used to quantify performance on the secondary task led to a similar pattern of results."

Desjardins, J.L. and Doherty, K.A., **2013**. Age-related changes in listening effort for various types of masker noises. Ear and hearing, 34(3), pp.261-272.
https://journals.lww.com/ear-hearing/fulltext/2013/05000/Age_Related_Changes_in_Listening_Effort_for.2.aspx?casa_token=-xKm8Y4coS0AAAAA:ydPRCMQhnFciaJ-kKHtuL56d0eTsXLUZlcswJtOqVzu8ryNb8bom2zjK9GKA0sHmJjHHVQ4u7Y6nc2wS_Q
Listening effort is how much time they cannot follow a ball with the mouse (visual motor tracking) ((Kieweg 2009)). "elliptical track with a circle-shaped target that rotates along the track and is displayed on a computer monitor." They modify the speed of the rotation around the ellipse (RPM). "baseline level of 80% time on target (TOT)"
"**Listening to and repeating** the sentences was done while continuously tracking the moving target on the computer screen."
"For each group of participants, there were no significant differences in listening effort measured across the masker conditions, with the exception of the younger participants, who expended more effort listening in the six-talker masker condition compared with the SSN condition."
"Akeroyd (2008), in a review of more than 20 studies examining speech recognition in noise and cognitive function, primary predictor of speech- recognition performance in noise was listeners' working-memory capacity, as measured by the Reading Span test."
Thanks "to Dr. Susan Kemper for the Digital Pursuit Rotor Tracking program."



Xia, J., Nooraei, N., Kalluri, S. and Edwards, B., **2015**. Spatial release of cognitive load measured in a dual-task paradigm in normal-hearing and hearing-impaired listeners. The Journal of the Acoustical Society of America, 137(4), pp.1888-1898.
https://asa.scitation.org/doi/abs/10.1121/1.4916599
Also use the ball (random pattern). They use the CRM task and **separate** by **listening and production** (stimulus and response). Contra-indicative results. I am still wondering if the listening effort was higher when speaking. According to the plots it is not, but it could be that they have two baselines, one for listening and one for speaking. "Tracking accuracy declined when the tracking task was done concurrently with the speech recognition task"
"Given the very poor recognition of the target speech, the low listening cost in the no-cue condition may indicate that participants were disengaged from the primary listening task and diverted all of their cognitive resources to the tracking task,"
"The participants initiated the tracking task by placing their fingers on a touch screen. During one tracking block, the visual target, a solid white dot 1 cm in diameter, was continuously moving around on the touch screen in a random pattern."
"The target movement speed was determined before experimental testing for each participant to achieve an on-target tracking accuracy of 60%.""The target speed was 1 pixel per frame (PPF) in the initial tracking block."



Kemper, S., Schmalzried, R., Herman, R., Leedahl, S. and Mohankumar, D., **2009**. The effects of aging and dual task demands on language production. Aging, Neuropsychology, and Cognition, 16(3), pp.241-259.

https://doi.org/10.1080/13825580802438868
Language production with the ball doing an ellipse. She is the creator of the DPRT (digital visual pursuit rotor tracking).
"Initial tracking speeds for young and older adults were approximately 2 and .75 revolutions per min, respectively" 3.4 rev/min young, 1.5 rev/min elder.
"Participants practiced tracking for 30 s and received feedback on their tracking performance.A “ 2 up/1 down stair-case” training procedure was used to gradually increase tracking speed on successive 30-s trials: if average time on target was 80% or better for a trial, the speed was increased by 10% for the next trial; if average time on target was less than 80%, the speed was decreased by 5%. The stair-case procedure converged on an asymptotic tracking speed when the speed oscillated around the same value, moving “up” and “down” past this value 3 times. " 23.8 trials of 30s for young to find this number.

Peng, Z.E. and Wang, L.M., **2019**. Listening effort by native and nonnative listeners due to noise, reverberation, and talker foreign accent during english speech perception. Journal of Speech, Language, and Hearing Research, 62(4), pp.1068-1081.
https://www.researchgate.net/publication/332216096_Listening_Effort_by_Native_and_Nonnative_Listeners_Due_to_Noise_Reverberation_and_Talker_Foreign_Accent_During_English_Speech_Perception
TO READ
"Performance on the APR task did not vary with changing acoustic conditions, but it did suggest increased listening effort for native listeners of other languages compared to the 2 other listener groups. " adaptive pursuit rotor (APR) Similar to Trisha's dual task.

QUESTIONS
Dua-tasks should take same areas/processes of the brain? For example, following a ball with a mouse on the screen is a psychomotor task and visual, thus not related to speech processing and decision making. These are different areas/processes of the brain right? Thus there should be little effect on the primary-secondary task performance. Is there any research connecting brain areas/processes and dual-tasks?

STATISTICS

Liddell, T.M. and Kruschke, J.K., **2018**. Analyzing ordinal data with metric models: What could possibly go wrong?. Journal of Experimental Social Psychology, 79, pp.328-348.
https://www.sciencedirect.com/science/article/pii/S0022103117307746
https://media.dlib.indiana.edu/media_objects/mp48sh50s
There is a talk online that explains the paper. Also references to easy-introductory papers about bayesian statistics.

Parker, R.I. and Vannest, K., **2009**. An improved effect size for single-case research: Nonoverlap of all pairs. Behavior therapy, 40(4), pp.357-367.
https://www.sciencedirect.com/science/article/pii/S0005789408000816

Reviewer: "The NAP effect measure (Parker & Vannest, 2009), used by, e.g., von Gablenz et al. (2021), is a point estimate of the probability that the respondent rates some perceptual quality higher in condition B than in condition A. Thus, the NAP is a measure on a well-defined interval scale for each respondent, and it uses only the ordinal property of the ratings, without assuming a metric scale.

I am not suggesting the authors should completely abandon their main analysis approach. I only suggest it would strengthen the paper if they also include a test of the main conclusion using, e.g., a Wilcoxon test on the 16 individual NAP values for the loudness difference between field and lab, in relation to the null hypothesis that NAP = 0.5."

NAP reports the significance for single users between two conditions. NAP was used to give a nonparametric ES (effect size) estimate for the HA intervention

von Gablenz, P., Kowalk, U., Bitzer, J., Meis, M. and Holube, I., **2021**. Individual hearing aid benefit in real life evaluated using ecological momentary assessment. Trends in hearing, 25, p.2331216521990288.
https://journals.sagepub.com/doi/full/10.1177/2331216521990288
Uses NAP.

Within-subject of ordinal variables
https://www.scalestatistics.com/friedmans-anova.html