Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect

Kristin J. Van Engen,
Departments of Linguistics and Communication Sciences and Disorders, The University of Texas at Austin

Zilong Xie, and
Department of Communication Sciences and Disorders, The University of Texas at Austin

Bharath Chandrasekaran
Department of Communication Sciences and Disorders, The University of Texas at Austin

Abstract

In noisy situations, visual information plays a critical role in the success of speech communication: listeners are better able to understand speech when they can see the speaker. Visual influence on auditory speech perception is also observed in the McGurk effect, in which discrepant visual information alters listeners' auditory perception of a spoken syllable. When hearing /ba/ while seeing a person saying /ga/, for example, listeners may report hearing /da/. Because these two phenomena have been assumed to arise from a common integration mechanism, the McGurk effect has often been used as a measure of audiovisual integration in speech perception. In this study, we test whether this assumed relationship exists within individual listeners. We measured participants' susceptibility to the McGurk illusion as well as their ability to identify sentences in noise across a range of signal-to-noise ratios (SNRs) in audio-only and audiovisual modalities. 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.

Keywords

audiovisual speech perception; speech perception in noise; McGurk effect

Introduction

A brief foray into the literature on audiovisual speech perception reveals a common rhetorical theme, in which authors begin with the general claim that visual information influences speech perception and cite two effects as evidence: the McGurk effect (McGurk and McDonald, 1976) and the intelligibility benefit garnered by listeners when they can see

speakers' faces (Sumby and Pollack, 1954). (For examples of papers that begin this way, see Altieri et al., 2011; Anderson et al., 2009; Colin et al., 2005; Grant et al., 1998; Magnotti et al., 2015; Massaro et al., 1993; Nahorna et al., 2012; Norrix et al., 2007; Ronquest et al., 2010; Rosenblum et al., 1997; Ross et al., 2007; Saalasti et al., 2011; Sams et al., 1998; Sekiyama, 1997; Sekiyama et al., 2003; Strand et al., 2014; van Wassenhove et al., 2007.) Both effects have been replicated many times and unquestionably show the influence of visual input on speech perception.

It is often assumed, then, that these two phenomena arise from a common audio-visual integration mechanism. As a result, the McGurk effect (i.e., auditory misperception of a spoken syllable when it is presented with incongruent visual information) has often been used as a measure of auditory-visual integration in speech perception. van Wassenhove, et al. (2007), for example, define AV speech integration as having occurred when "a unitary integrated percept emerges as the result of the integration of *clearly differing* auditory and visual informational content", and therefore use the McGurk illusion to "quantify the degree of integration that has taken place." (p. 598). Alsius et al. (2007) similarly define the degree of audiovisual integration as "the prevalence of the McGurk effect." (p.400). Moreover, a number of studies investigating audiovisual speech perception in clinical populations (e.g., individuals with schizophrenia (Pearl et al., 2009), children with amblyopia (Burgmeier et al., 2015), and individuals with Asperger Syndrome (Saalasti et al., 2011)) have used the McGurk effect as their primary dependent measure of audiovisual speech processing.

Despite the popularity of this approach, it still remains to be convincingly demonstrated that an individual's susceptibility to the McGurk illusion relates to their ability to take advantage of visual information during everyday speech processing. There is evidence that McGurk susceptibility relates (weakly) to lipreading ability under some task and scoring conditions (Strand et al., 2014), with better lipreaders being slightly more susceptible to the McGurk effect. There is also evidence linking lipreading ability to audiovisual speech perception (Grant et al., 1998). The connection between McGurk susceptibility and audiovisual speech perception was investigated in one study on older adults with acquired hearing loss (Grant and Seitz, 1998), with equivocal results: there was a correlation between McGurk susceptibility and visual enhancement for sentence recognition (r=.46), but McGurk susceptibility did not contribute significantly to a regression model predicting visual enhancement. The relationship between McGurk susceptibility and the use of visual information during speech perception, therefore, remains unclear. Here we present a within-subjects study of young adults with normal hearing in which we assess McGurk susceptibility and audiovisual sentence recognition across a range of noise levels and types. If susceptibility to the McGurk effect reflects an audiovisual integration process that is relevant to everyday speech comprehension in noise, then we expect listeners who are more susceptible to the illusion to show greater speech intelligibility gains when visual information is available for sentence recognition. If, on the other hand, different mechanisms mediate the use of auditory and visual information in the McGurk task and during everyday speech perception, then no such relationship is predicted. Such a finding would cast doubt on the utility of the McGurk task as a measure of audiovisual speech perception.

Method

Participants

39 healthy young adults (18 to 29 years; mean age = 21.03 years) were recruited from the Austin, Texas community. All participants were native speakers of American English and reported no history of speech, language or hearing problems. Their hearing was screened to ensure thresholds 25 dB HL at 1000, 2000, and 4000 Hz for each ear, and their vision was normal or corrected-to-normal. Participants were compensated in accordance with a protocol approved by the University of Texas Institutional Review Board.

McGurk task

Stimuli—The stimuli were identical to those in Experiment 2 of Mallick et al. (2015). They consisted of two types of AV syllables: McGurk incongruent syllables (auditory /ba/ + visual /ga/) and congruent syllables (/ba/, /da/, and /ga/). The McGurk syllables were created using video recordings from eight native English speakers (4 females, 4 males). A different female speaker recorded the three congruent syllables.[1]

Procedure—The task was administered using E-Prime 2.0 software (Schneider et al. 2002). Auditory stimuli were presented binaurally at a comfortable level using Sennheiser HD-280 Pro headphones, and visual stimuli were presented on a computer screen. A fixation cross was displayed for 500 ms prior to each stimulus. Following Mallick et al., (2015), participants were instructed to report the syllable they heard in each trial from the set /ba/, /da/, /ga/, and /tha/. The eleven stimuli (eight McGurk and three congruent) were each presented ten times. The presentation of these 110 stimuli was randomized and self-paced.

McGurk susceptibility—Responses to the McGurk incongruent stimuli were used to measure listeners' susceptibility to the McGurk effect. As in Mallick et al (2015), responses of either /da/ or /tha/ were coded as McGurk fusion percepts.

Speech perception in noise task

Target speech stimuli—A young adult male speaker of American English produced 80 simple sentences, each containing four keywords (e.g., *The gray mouse ate the cheese.*) (Van Engen et al., 2012). Sentences were used for this task (rather than syllables) because our interest is in the relationship between McGurk susceptibility and the processing of running speech.

Maskers—Two maskers, equated for RMS amplitude, were generated to create speech-in-noise stimuli: speech-shaped noise (SSN) filtered to match the long-term average spectrum of the target speech and two-talker babble consisting of two male voices. The two maskers were included to assess listeners' ability to take advantage of visual cues in different types of challenging listening environments. SSN renders portions of the target speech signal inaudible to listeners (i.e., energetic masking), while two-talker babble can also interfere

---

[1]The unimodal intelligibility of these stimuli were tested by Mallick et al. (2015). Auditory stimuli were identified with 97% accuracy (SD = 4%) and visual stimuli were identified with 80% accuracy (SD=10%).

with target speech identification by creating confusion and/or distraction not accounted for by the physical properties of the speech and noise (i.e., informational masking). Visual information is more helpful to listeners when the masker is composed of other voices (Helfer and Freyman, 2005).

Mixing targets and maskers—The audio was detached from the video recording of each sentence and RMS amplitude equalized using Praat (Boersma and Weenink 2010). Each audio clip was mixed with the maskers at 5 levels to create stimuli with the following signal-to-noise ratios (SNRs): −4 dB, −8 dB, −12 dB, −16 dB, and −20 dB. Each noise clip was 1s longer than its corresponding sentence so that 500 ms of noise could be played before and after each target. These mixed audio clips served as the stimuli for the audio-only (AO) condition. The audio files were also reattached to the corresponding videos to create the stimuli for the audiovisual (AV) condition. In total, there were 400 audio files and 400 corresponding audiovisual files with the SSN masker (80 sentences × 5 SNRs), and 400 final audio files and 400 corresponding audiovisual files with the two-talker babble masker (80 sentences × 5 SNRs).

Design and Procedure—Masker type (SSN and two-talker babble), modality (AO and AV), and SNR (−4 dB, −8 dB, −12 dB, −16 dB, and −20 dB) were manipulated within subjects. Four target sentences were presented in each condition for a total of 80 trials. Trials were randomized for each participant. No sentence was repeated for a given participant.

Stimuli were presented to listeners at a comfortable level. Participants were instructed that they would be listening to AO and AV sentences in noise, and were told that the target sentences would begin a half-second after the noise. The participant initiated each stimulus presentation using the keyboard, and they were given unlimited time to respond. If they were unable to understand a sentence, they were asked to report any intelligible words and/or make their best guess. If they did not understand anything, they were told to type 'X'. For AO trials, a centered crosshair was presented on the screen during the audio stimulus; for AV trials, a full-screen video of the speaker was presented with the audio. Responses were scored by the number of keywords identified correctly. Homophones and obvious spelling errors were scored as correct; words with added or deleted morphemes were scored as incorrect.

Results

McGurk susceptibility

Figure 1 shows the distribution of McGurk susceptibility scores. As in Mallick et al. (2015), scores ranged from 0–100% and were skewed to the right.

Keyword identification in noise

Average keyword intelligibility across SNRs is shown in Figure 2. The full set of identification data was first analyzed to determine whether McGurk susceptibility predicted keyword identification under any of the test conditions. Statistical analysis was performed using the lme4 package (version 1.1–12, Bates et al., 2015) for R software (2016). The

response to each keyword was categorized as correct or incorrect and analyzed using a mixed logit model for binomially distributed outcomes. Because comparing the noise types to one another was not of primary interest, separate analyses were conducted for the two noise types to simplify model interpretation. For each analysis, modality, SNR, McGurk scores and their 2- and 3-way interactions were entered as fixed factors. Modality was deviation-coded (i.e., −0.5 and 0.5), which entails that the coefficient represents modality's "main effect" (i.e., its partial effect when all others are zero). Continuous predictors (SNR and McGurk scores) were centered and scaled. The models were fit with the maximal random effects structure justified by the experimental design (Barr et al., 2013)[2]. The model outputs for the fixed effects are shown in Tables 1–2.

For both types of noise, modality, SNR, and their interaction were the only statistically significant predictors of keyword identification in noise. Neither McGurk susceptibility nor its interactions with SNR or modality significantly predicted keyword identification. That said, McGurk scores and their interaction with modality did near statistical significance (p=. 10, p=.07) in SSN. As shown in Figure 3, these trends reflect a negative association between McGurk susceptibility, particularly in the AV conditions.

Visual enhancement

One way researchers quantify individuals' ability to use visual information during speech identification tasks is by calculating *visual enhancement* (VE), which takes the difference between a listener's performance in AV and AO conditions and normalizes it by the proportion of improvement available given their AO performance (Grant and Seitz, 1998; Grant et al., 1998; Sommers et al., 2005):

$$VE = \frac{(AV - AO)}{(1 - AO)}$$

Because AO performance must be below ceiling to calculate VE (i.e., there must be room for improvement), VE was calculated for the SNRs of −8 dB and below. (At least 10% of the subjects were at ceiling in AO at −4 dB). For each listener, VE was calculated separately for 2-talker babble and SSN. A positive VE score indicates that a listener identified more words in the AV condition than in the AO condition; the maximum VE score of 1 indicates that the listener identified all of the keywords in the AV condition.

Results are shown in Figure 4. One outlier is not displayed because it would have required a significant extension of the y-axis. (In SSN at −8 dB, the individual identified ~94% of AO words, but only 50% of the AV words, resulting in a VE score of −7.) All other data are shown. Although it is not the focus of this study, it is worth noting that the VE data do not follow the principle of inverse effectiveness (Holmes 2009), which predicts greater multisensory integration in more difficult conditions. If anything, the data for SSN suggest the opposite pattern: greater VE at easier SNRs. (See also Van Engen et al. (2014), Tye-

---

[2]For this study, the maximal random effects structure included random intercepts for subjects and sentences and random slopes for the following: subject by modality, subject by SNR, subject by the interaction of modality and SNR, sentence by modality, sentence by SNR, and sentence by the interaction of modality and SNR.

Murray et al. (2010), and Ross et al. (2007) for other cases where this principle does not capture behavior in AV speech perception).

Figure 5 displays the data for each condition with McGurk susceptibility scores plotted against VE scores. Due to the skewness of the McGurk data, Kendall's tau was used to assess the relationship between individuals' rate of McGurk responses and their VE scores. As shown in Table 3, there was no statistically significant association in any condition (p-values ranged from .06 to .85; tau values ranged from −.21 to .06).

In SSN at −8 dB and in 2-talker babble at −16 dB, this correlation neared statistical significance (p=.190 and p=.063) with negative tau values, indicating that the (possible) relationship between VE and McGurk susceptibility is one in which susceptibility to the illusion is associated with *less* VE. In both the analysis of keyword identification and VE, then, no association between McGurk susceptibility and speech perception in noise was significant at the p <.05 level, and the two marginally significant relationships suggested that susceptibility to the illusion was associated with lower rates of keyword identification and less visual enhancement.

Discussion

The results of this experiment showed no statistically significant relationship between an individual's susceptibility to the McGurk effect and their ability to understand speech in noise (with or without visual information). Importantly, the VE analyses also showed no significant relationship between an individual's susceptibility to the McGurk illusion and their ability to use visual cues in order to improve on audio-only speech perception.[3] Research that is fundamentally interested in understanding audiovisual integration as it relates to a listener's ability to understand connected speech, therefore, should not assume that susceptibility to the McGurk effect can be used as a valid measure of AV speech processing.

Although the McGurk effect and the benefit of visual information both show that visual information affects auditory processing, there are several critical differences between these phenomena. One noteworthy difference is the congruency of the auditory and visual information in these situations. In the McGurk effect, auditory identification of individual consonants is altered by conflicting visual information; visual information can be thought of, therefore, as detrimental to correct auditory perception. In AO versus AV speech in noise, *congruent* visual information *facilitates* auditory perception. At least one recent neuroimaging study supports the hypothesis that different neural mechanisms may mediate the integration of congruent versus incongruent visual information with auditory signals. Using fMRI, Erickson et al. (2014) showed that distinct posterior superior temporal regions are involved in processing congruent AV speech and incongruent AV speech when compared

---

[3]This result contrasts with that of Grant and Seitz (1998), which showed a significant correlation between McGurk susceptibility and visual enhancement for sentence recognition in older adults with hearing loss. (Note, however, that the regression analysis in that study also indicated McGurk susceptibility was not a significant predictor of VE). Given the different stimulus materials, noise levels, and listener populations in the two studies, there are multiple possible explanations for the different outcomes. Differences in unimodal abilities across the participants in the two studies are a likely candidate: not only did the older adults in Grant and Seitz (1998) have hearing loss, but speechreading ability is also known to decline with age (Sommers et al., 2005; Tye-Murray et al., 2016).

to unimodal speech (acoustic-only and visual-only). Left posterior superior temporal sulcus (pSTS) was recruited during congruent bimodal AV speech. In contrast, left posterior superior temporal gyrus (pSTG) was recruited when processing McGurk speech, suggesting that left pSTG may be necessary when there is a discrepancy between auditory and visual cues. It may be that left pSTG is involved in the generation of the fused percept that can arise from conflicting cues.

Another critical difference between McGurk tasks and sentence-in-noise tasks is the amount of linguistic context available to listeners. The top-down and bottom-up processes involved in identifying speech vary significantly with listeners' access to lexical, syntactic, and semantic context (Mattys et al., 2005), and the availability of rhythmic information in running speech allows for neural entrainment and predictive processing that is not possible when identifying isolated syllables (Peelle and Davis, 2012; Peelle and Sommers, 2015). In keeping with these observations, previous studies have failed to show a relationship between AV integration measures derived from consonant versus sentence recognition (Grant and Seitz, 1998; Sommers et al., 2005). Grant and Seitz (1998) showed no correlation between the two in older adults with hearing impairment, and Sommers et al. (2005) found that, while VE for word and sentence identification were related to one another for young adults, VE for consonant identification was not related to either. Given that other researchers *have* shown significant relationships between consonant identification and words or sentences in *unimodal* conditions (Humes et al., 1994; Grant et al., 1998; Sommers et al. 2005), these results suggest that the lack of relationship between VE for consonants and VE for words and sentences results from differences in the mechanisms mediating the integration of auditory and visual inputs for these different types of speech materials.

The results reported here serve two purposes: first, as a caution against the assumption that the McGurk effect can be used as an assay of audiovisual integration for speech in challenging listening conditions; and second, to set the scene for future work investigating the potentially different mechanisms supporting the integration of auditory and visual information across different types of speech materials.

Acknowledgments

References

Altieri N, Pisoni DB, Townsend JT. Some behavioral and neurobiological constraints on theories of audiovisual speech integration: a review and suggestions for new directions. Seeing and Perceiving. 2011; 24(6):513–539. http://doi.org/10.1163/187847611X595864. [PubMed: 21968081]

Alsius A, Navarra J, Soto-Faraco S. Attention to touch weakens audiovisual speech integration. Experimental Brain Research. 2007; 183(3):399–404. http://doi.org/10.1007/s00221-007-1110-1. [PubMed: 17899043]

Andersen TS, Tiippana K, Laarni J, Kojo I, Sams M. The role of visual spatial attention in audiovisual speech perception. Speech Communication. 2009; 51(2):184–193. http://doi.org/10.1016/j.specom.2008.07.004.

Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language. 2013; 68(3):255–278. http://doi.org/10.1016/j.jml.2012.11.001.

Bates D, Mächler M, Bolker BM, Walker SC. Fitting Linear Mixed-Effects Models using lme4. Journal of Statistical Software. 2015; 67(1):1–48.

Boersma, P., Weenink, D. Praat: doing phonetics by computer (Version 5.1). 2010. Retrieved from http://www.praat.org

Burgermeier R, Desai RU, Farner KC, Tiano B, Lacey R, Volpe NJ, Mets MB. The Effect of Amblyopia on Visual-Auditory Speech Perception. 2015; 133(1):11. http://doi.org/10.1001/jamaophthalmol.2014.3307.

Colin C, Radeau M, Deltenre P. Top-down and bottom-up modulation of audiovisual integration in speech. European Journal of Cognitive Psychology. 2005; 17(4):541–560. http://doi.org/10.1080/09541440440000168.

Erickson LC, Zielinski BA, Zielinski JEV, Liu G, Turkeltaub PE, Leaver AM, Rauschecker JP. Distinct cortical locations for integration of audiovisual speech and the McGurk effect. Frontiers in Psychology. 2014; 5(158):265. http://doi.org/10.3389/fpsyg.2014.00534. [PubMed: 24734026]

Grant KW, Walden BE, Seitz PF. Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. The Journal of the Acoustical Society of America. 1998; 103(5):2677–2690. http://doi.org/10.1121/1.422788. [PubMed: 9604361]

Grant KW, Seitz PF. Measures of auditory–visual integration in nonsense syllables and sentences. The Journal of the Acoustical Society of America. 1998; 104(4):2438–2450. http://doi.org/10.1121/1.423751. [PubMed: 10491705]

Helfer KS, Freyman RL. The role of visual speech cues in reducing energetic and informational masking. The Journal of the Acoustical Society of America. 2005; 117(2):842–849. http://doi.org/10.1121/1.1836832. [PubMed: 15759704]

Holmes NP. The Principle of Inverse Effectiveness in Multisensory Integration: Some Statistical Considerations. Brain Topography. 2009; 21(3–4):168–176. http://doi.org/10.1007/s10548-009-0097-2. [PubMed: 19404728]

Humes LE, Watson BU, Christensen LA, Cokely CG, Halling DC, Lee L. Factors associated with individual differences in clinical measures of speech recognition among the elderly. Journal of Speech Language and Hearing Research. 1994; 37(2):465–474. http://doi.org/10.1044/jshr.3702.465.

Mallick DB, Magnotti JF, Beauchamp MS. Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type. Psychonomic Bulletin & Review. 2015; 22(5):1299–1307. http://doi.org/10.3758/s13423-015-0817-4. [PubMed: 25802068]

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264(5588):746–748. http://doi.org/10.1038/264746a0. [PubMed: 1012311]

Magnotti JF, Mallick DB, Feng G, Zhou Bin, Zhou W, Beauchamp MS. Similar frequency of the McGurk effect in large samples of native Mandarin Chinese and American English speakers. Experimental Brain Research. 2015; 233(9):2581–2586. http://doi.org/10.1007/s00221-015-4324-7. [PubMed: 26041554]

Massaro DW, Cohen MM, Gesi A, Heredia R, Tsuzaki M. Bimodal speech perception: an examination across languages. Journal of Phonetics. 1993; 21:445–478.

Mattys SL, White L, Melhorn JF. Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 134(4):477–500. http://doi.org/10.1037/0096-3445.134.4.477.

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264(5588):746–748. http://doi.org/10.1038/264746a0. [PubMed: 1012311]

Nahorna O, Berthommier F. Binding and unbinding the auditory and visual streams in the McGurk effect. The Journal of the Acoustical Society of America. 2012:1061–1077. http://doi.org/10.1121/1.4728187. [PubMed: 22894226]

Nath AR, Beauchamp MS. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. NeuroImage. 2012; 59(1):781–787. http://doi.org/10.1016/j.neuroimage.2011.07.024. [PubMed: 21787869]

Norrix LW, Plante E, Vance R, Boliek CA. Auditory-Visual Integration for Speech by Children With and Without Specific Language Impairment. Journal of Speech Language and Hearing Research. 2007; 50(6):1639–1651. http://doi.org/10.1044/1092-4388(2007/111).

Pearl D, Yodashkin-Porat D, Katz N, Valevski A, Aizenberg D, Sigler M, et al. Differences in audiovisual integration, as measured by McGurk phenomenon, among adult and adolescent patients with schizophrenia and age-matched healthy control groups. Comprehensive Psychiatry. 2009; 50(2):186–192. http://doi.org/10.1016/j.comppsych.2008.06.004. [PubMed: 19216897]

Peelle JE, Davis MH. Neural oscillations carry speech rhythm through to comprehension. Frontiers in Psychology. 2012; 3 http://doi.org/10.3389/fpsyg.2012.00320.

Peelle JE, Sommers MS. Prediction and constraint in audiovisual speech perception. Cortex. 2015; 68:169–181. http://doi.org/10.1016/j.cortex.2015.03.006. [PubMed: 25890390]

R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016. URL https://www.R-project.org/

Ronquest RE, Levi SV, Pisoni DB. Language identification from visual-only speech signals. Attention, Perception, & Psychophysics. 2010; 72(6):1601–1613. http://doi.org/10.3758/app.72.6.1601.

Rosenblum LD, Schmuckler MA, Johnson JA. The McGurk effect in infants. Perception and Psychophysics. 1997; 59(3):347–357. http://doi.org/10.3758/BF03211902. [PubMed: 9136265]

Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cerebral Cortex. 2007; 17(5):1147–1153. http://doi.org/10.1093/cercor/bhl024. [PubMed: 16785256]

Saalasti S, Kätsyri J, Tiippana K, Laine-Hernandez M, Wendt, von L, Sams M. Audiovisual Speech Perception and Eye Gaze Behavior of Adults with Asperger Syndrome. Journal of Autism and Developmental Disorders. 2011; 42(8):1606–1615. http://doi.org/10.1007/s10803-011-1400-0.

Sams M, Manninen P, Surakka V, Helin P. McGurk effect in Finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context. Speech Communication. 1998; 26(1–2):75–87. http://doi.org/10.1016/s0167-6393(98)00051-x.

Schneider, W., Eschman, A., Zuccolotto, A. E-Prime User's Guide. Pittsburgh: Psychology Software Tools, Inc.; 2002.

Sekiyama K. Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. Perception and Psychophysics. 1997; 59(1):73–80. http://doi.org/10.3758/BF03206849. [PubMed: 9038409]

Sekiyama K, Kanno I, Miura S, Sugita Y. Auditory-visual speech perception examined by fMRI and PET. Neuroscience Research. 2003; 47(3):277–287. http://doi.org/10.1016/S0168-0102(03)00214-1. [PubMed: 14568109]

Sommers MS, Tye-Murray N, Spehar B. Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. Ear and Hearing. 2005; 26(3):263–275. http://doi.org/10.1097/00003446-200506000-00003. [PubMed: 15937408]

Strand J, Cooperman A, Rowe J, Simenstad A. Individual differences in susceptibility to the McGurk effect: links with lipreading and detecting audiovisual incongruity. Journal of Speech Language and Hearing Research. 2014; 57(6):2322–2331. http://doi.org/10.1044/2014_JSLHR-H-14-0059.

Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. The Journal of the Acoustical Society of America. 1954; 26(2):212–215. http://doi.org/10.1121/1.1907309.

Tiippana K. What is the McGurk effect? Frontiers in Psychology. 2014; 5 http://doi.org/10.3389/fpsyg.2014.00725.

Tye-Murray N, Sommers M, Spehar B, Myerson J. Aging, audiovisual integration, and the principle of inverse effectiveness. Ear and Hearing. 2010:1. http://doi.org/10.1097/aud.0b013e3181ddf7ff.

Tye-Murray N, Spehar B, Myerson J, Hale S, Sommers M. Lipreading and audiovisual speech recognition across the adult lifespan: implications for audiovisual integration. Psychology and Aging. 2016; 31(4):380–389. http://dx.doi.org/10.1037/pag0000094. [PubMed: 27294718]

Van Engen KJ, Chandrasekaran B, Smiljanic R. Effects of speech clarity on recognition memory for spoken sentences. PloS One. 2012; 7(9):e43753. http://doi.org/10.1371/journal.pone.0043753. [PubMed: 22970141]

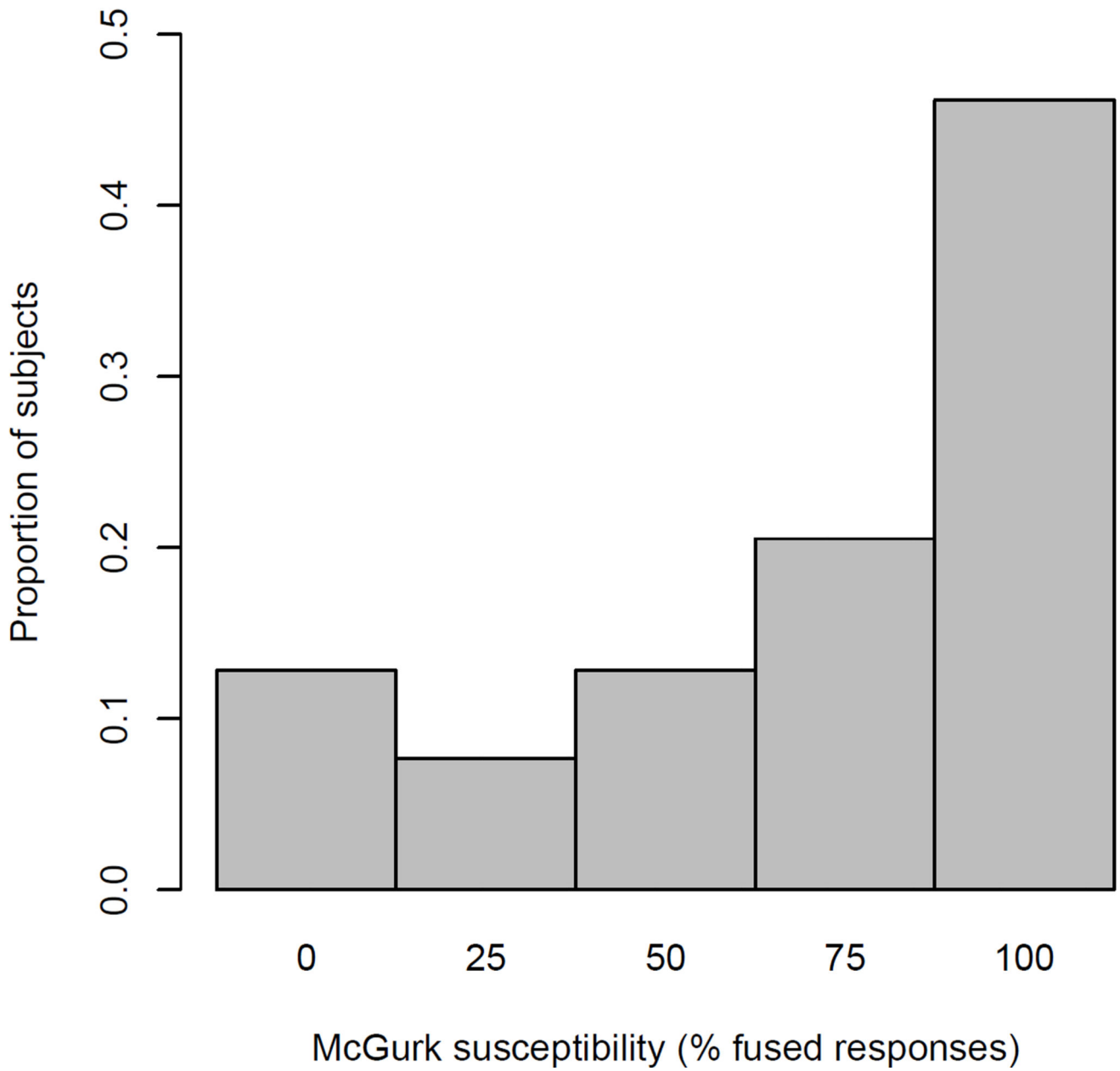van Wassenhove V, Grant KW, Poeppel D. Temporal window of integration in auditory-visual speech perception. Neuropsychologia. 2007; 45(3):598–607. http://doi.org/10.1016/j.neuropsychologia.2006.01.001. [PubMed: 16530232]

Distribution of McGurk susceptibiliy scores

Figure 1.
Distribution of McGurk susceptibility scores

Figure 2.
Proportion of keywords identified across SNRs in speech-shaped noise (left) and 2-talker babble (right). Error bars represent standard error.
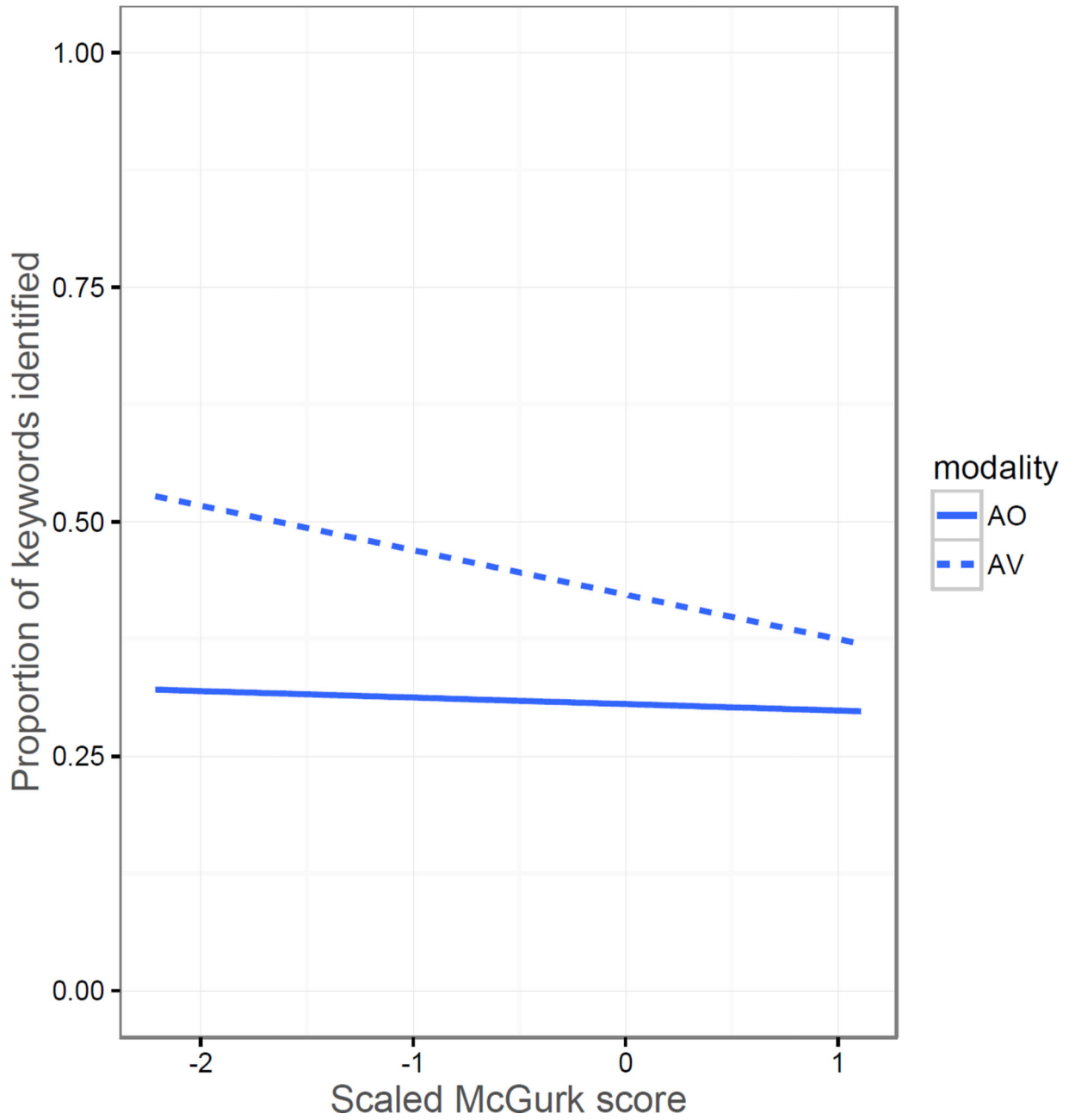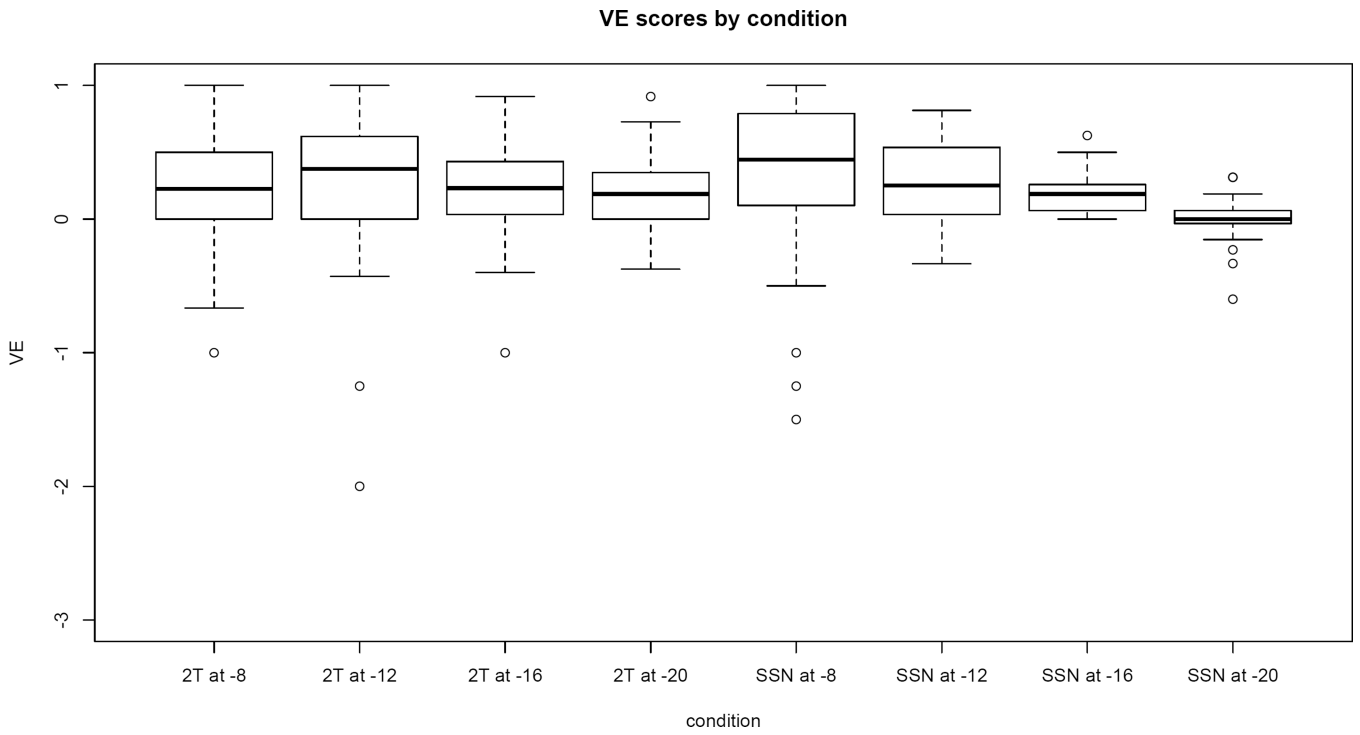
Keyword identification x McGurk susceptibility

Figure 3.
Relationship between McGurk susceptibility scores (scaled) and proportion of keywords identified in SSN.

Figure 4.

VE scores by condition. The boxes extend from the 25th percentile to the 75th percentile, with the dark line indicating the median. Whiskers extend to data points that are within 1.5 times the interquartile range. Data outside that range is denoted by open circles. Note that one outlier is not displayed. (A VE score of −7 on SSN at −8 dB)
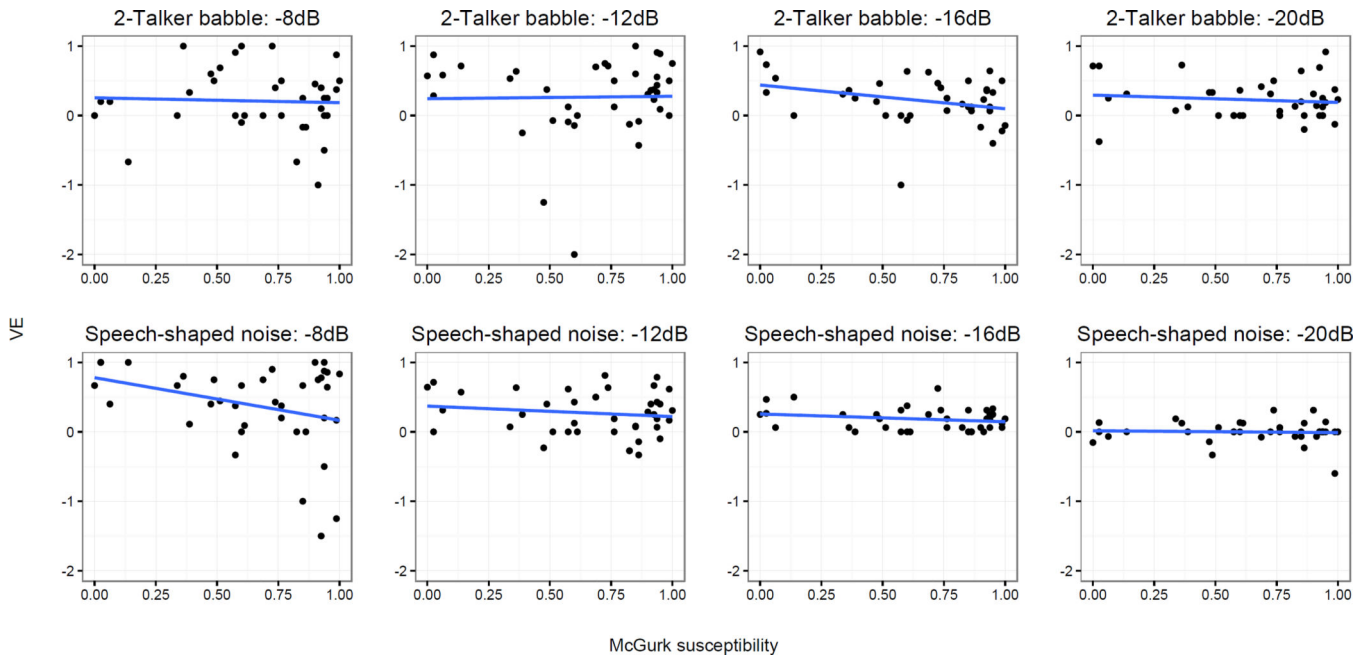
Figure 5.

Visual enhancement plotted against McGurk susceptibility for each listening condition. Top row: 2-talker babble conditions. Bottom row: Speech-shaped noise conditions. Linear trendlines are included, with shading to represent standard error. For the sake of axis consistency, one data point for SSN −8dB is not displayed, although it is included in the calculation of the trendline (McGurk score=.86; VE score =−7).

Table 1

Speech-shaped noise

Fixed effects	Estimate	SE	Z value	P-value
Intercept	−1.41497	0.20652	−6.851	<0.001
Modality	**1.58603**	**0.27298**	**5.810**	**<0.001**
SNR	**−2.78651**	**0.15256**	**−18.266**	**<0.001**
McGurk	−0.29178	0.17577	−1.660	0.09690
Modality × SNR	**0.79314**	**0.30170**	**2.629**	**0.00857**
Modality × McGurk	−0.37224	0.20761	−1.793	0.07297
SNR × McGurk	0.03589	0.10334	0.347	0.72836
Modality × SNR × McGurk	0.21852	0.19936	1.096	0.27303

Table 2

Two-talker babble

Fixed effects	Estimate	SE	Z value	P-value
Intercept	−0.67487	0.28097	−2.402	0.0163
Modality	**1.37872**	**0.20354**	**6.774**	**<0.001**
SNR	**−1.56604**	**0.12803**	**−12.231**	**<0.001**
McGurk	−0.21197	0.25670	−0.826	0.4090
Modality × SNR	**0.80622**	**0.17795**	**4.531**	**<0.001**
Modality × McGurk	−0.15580	0.15740	−0.990	0.3223
SNR × McGurk	0.06916	0.09763	0.708	0.4787
Modality × SNR × McGurk	−0.16874	0.10971	−1.538	0.1240

Table 3

	z	p-value	tau
SSN: −8 dB	−1.312	0.190	−0.150
SSN: −12 dB	−0.376	0.707	−0.043
SSN: −16 dB	−.676	0.499	−0.080
SSN: −20 dB	−0.192	0.848	−0.023
2T: −8 dB	−0.261	0.794	−0.031
2T: −12 dB	0.533	0.594	0.060
2T: −16 dB	−1.855	0.063	−0.210
2T: −20 dB	−0.612	0.541	−0.071