Ventriloquist Illusion Produced With Virtual Acoustic Spatial Cues and Asynchronous Audiovisual Stimuli in Both Young and Older Individuals

Marnix Stawicki [1,2], **Piotr Majdak** [3] **and Deniz Başkent** [1,2,*]

[1] Department of Otorhinolaryngology / Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
[2] Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences (BCN), University of Groningen, Groningen, The Netherlands
[3] Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria

Abstract

Ventriloquist illusion, the change in perceived location of an auditory stimulus when a synchronously presented but spatially discordant visual stimulus is added, has been previously shown in young healthy populations to be a robust paradigm that mainly relies on automatic processes. Here, we propose ventriloquist illusion as a potential simple test to assess audiovisual (AV) integration in young and older individuals. We used a modified version of the illusion paradigm that was adaptive, nearly bias-free, relied on binaural stimulus representation using generic head-related transfer functions (HRTFs) instead of multiple loudspeakers, and tested with synchronous and asynchronous presentation of AV stimuli (both tone and speech). The minimum audible angle (MAA), the smallest perceptible difference in angle between two sound sources, was compared with or without the visual stimuli in young and older adults with no or minimal sensory deficits. The illusion effect, measured by means of MAAs implemented with HRTFs, was observed with both synchronous and asynchronous visual stimulus, but only with tone and not speech stimulus. The patterns were similar between young and older individuals, indicating the versatility of the modified ventriloquist illusion paradigm.

---

* To whom correspondence should be addressed. E-mail: d.baskent@umcg.nl, d.baskent@rug.nl

1. Introduction

In daily life, perception often relies on integration of signals from multiple senses (see Note 1) (Beauchamp, 2005; de Gelder and Bertelson, 2003; Ernst and Bülthoff, 2004; Lovelace *et al*., 2003; Stein and Meredith, 1993, and for a more recent review, Chen and Vroomen, 2013). An example of such multi-sensory integration of auditory and visual stimuli is the ventriloquist illusion, the mis-localization of the source of an auditory stimulus (such as the ac-tual talker) when it is presented with a temporally synchronous but spatially discordant visual stimulus (such as the moving mouth of a puppet; for a re-view, see Vroomen and de Gelder, 2004). The ventriloquism effect has often been quantified using the spatial ventriloquist paradigm, in which the loca-tion of the perceived event for synchronously presented but spatially separated auditory and visual stimuli is reported, and compared to the location of the auditory stimulus presented alone (Bermant and Welch, 1976; Bertelson and Aschersleben, 1998). In this paradigm, the perceived target auditory position is pulled towards the visual stimulus (Pick *et al*., 1969) and a larger lateral displacement in reporting location is required for a correct judgment. This pulling yields an increase in the measured threshold in location compared to the auditory-only condition. Interestingly, this procedure is similar to the es-timation of the minimum audible angle (MAA), the discrimination threshold of two spatially discordant sound events, in other words, the smallest angle that a listener can distinguish between two spatially separated sound sources (Perrott and Saberi, 1990). Thus, when quantifying the amount of audiovisual (AV) interaction in terms of the size of the ventriloquism effect, we essentially could make use of the magnitude of the MAA.

Previous studies indicated the ventriloquist illusion to be a robust and near-optimal bimodal sensory integration (Alais and Burr, 2004), as the illusion could be induced even when individuals were instructed and trained to ignore (Vroomen *et al*., 1998) or to not attend to the visual stimulus (Vroomen *et al*., 2001). These observations combined implied the ventriloquist illusion to mainly rely on automatic sensory processes (Bertelson *et al*., 2000).

Due to this robustness and its automatic nature, as well as the relatively simple task involved, the ventriloquist illusion could be a potentially useful tool in characterizing AV integration in a variety of populations, for exam-ple, in identifying effects of aging, as well as age-related hearing loss. AV integration and aging has long been an interest for research and clinical pur-poses, but the relevant studies have produced at times mixed results. With aging, as a result of age-related sensory and cognitive changes, perception of speech becomes challenging, especially in noisy situations or when other forms of distortions (e.g., reverberation) are involved (e.g., Bergman *et al*., 1976). Visual cues can help improve speech perception (Hoffman *et al*., 2012;

Pichora-Fuller *et al*., 1995). It was hypothesized that older individuals may rely more on visual speech cues and show an enhanced AV integration to compensate for age-related sensory and cognitive changes (Cienkowski and Carney, 2002; de Boer-Schellekens and Vroomen, 2014; Freiherr *et al*., 2013; Laurienti *et al*., 2006; Tye-Murray *et al*., 2007). However, while some studies supported such superior AV integration with older individuals (e.g., Başkent and Bazo, 2011; Helfer, 1998; Laurienti *et al*., 2006) some others showed a smaller AV benefit in older individuals (Musacchia *et al*., 2009; Tye-Murray *et al*., 2008, 2010). Another argument for potentially increased AV integration with aging came from the difficulty by older individuals in inhibiting information from one sensory modality for a task conducted in another modality, resulting in an inherently stronger multisensory integration (e.g., Couth *et al*., 2018). Lastly, an increase in temporal integration was suggested in older individuals, likely caused by factors such as general cognitive slowing down or reduced early sensory memories (Fogerty *et al*., 2016; Fozard, 1990; Salthouse, 1996). While it is not yet fully understood how the age-related changes in temporal integration are mediated by the auditory and visual modalities (Saija *et al*., 2019), if the stimulus contributed from each modality varies with age, producing a fused AV percept may become challenging. Opposing this view, increased temporal integration may contribute to a longer time window where the auditory and visual inputs are fused into one AV object. Yet, support from literature has, again, been mixed for this idea, with some studies showing evidence for a longer temporal AV integration window with older individuals while some others showing no such evidence (Alm and Behne, 2013; Başkent and Bazo, 2011; Diederich *et al*., 2008; Hay-McCutcheon, 2009).

A number of factors in these studies may have complicated the interpretation of the findings. In some studies, the baseline auditory-only performance differed between young and older groups; baseline speech intelligibility was lower and baseline response times were longer with the older group. Hence, tasks relying on speech intelligibility or lipreading might have been affected by age-related sensory and cognitive changes (Pichora-Fuller *et al*., 1995; Saija *et al*., 2014; Sommers *et al*., 2005), complicating the investigation of an age-only effect on AV integration. The differing baselines can prevent a fair between-subject comparison of relative improvement in performance with addition of multisensory cues because of the so-called inverse effectiveness, i.e., the stronger effect from addition of stimuli conveyed via other senses as the effectiveness of the uni-sensory stimuli is low (Couth *et al*., 2018; Holmes, 2009; Laurienti *et al*., 2006; Stein and Meredith, 1993). A simpler task able to produce similar auditory-only performance across the subject groups may be advantageous in assessing the relative changes in performance as a result of AV integration of added visual stimuli.

The ventriloquist illusion does not necessarily rely on speech understanding — in fact, it can be conducted with much simpler auditory stimuli — and being a mostly an automatic process, it may minimize the potential confounds discussed above and provide a useful tool to explore age effects on AV integration. While earlier studies implied that auditory-only MAAs can be affected by aging (Strouse *et al.*, 1998), more recently, Otte *et al.* (2013) showed that MAAs in azimuth were relatively insensitive to it. An age-insensitive measure, such as MAA, would hence be expected to produce similar uni-sensory baseline performance between the young and older groups, minimizing the confound of inverse effectiveness.

Hence, in this study, as a first step, we explored the applicability and robustness of the illusion. More specifically, we used a modified version of the ventriloquist illusion measure that (1) relied on MAAs, (2) was reduced in response bias by measuring left/right judgments of the AV event in an interleaved adaptive staircase procedure (Bertelson and Aschersleben, 1998), (3) was adapted from the free-field procedure (Bertelson and Aschersleben, 1998) to binaural stimulus reproduction via headphones (Wightman and Kistler, 1989), by using generic and easily available head-related transfer functions (HRTFs; Shaw, 1974), (4) used both non-speech (tones) and speech (words) stimuli, as these differ in stimulus complexity and related perceptual mechanisms, likely inducing differences in the AV integration processes (Lalonde and Holt, 2016; Tuomainen *et al.*, 2005), (5) was tested with both young and (with nearly normal hearing) older individuals and with stimuli adjusted to minimize further potential age-related hearing-loss effects, and (6) used both synchronous and asynchronous A and V stimuli as the temporal synchronicity may modulate the AV integration differently for younger and older individuals. We expected that, if the ventriloquist illusion is robust, our modified paradigm, combined with matching baseline auditory-only performance, would provide a useful tool that is simple to implement and easy to use to systematically investigate AV integration in young and older populations.

2. Materials and Methods

2.1. Participants

Two groups of native-Dutch speakers, young and older, participated in this study. The inclusion criteria were normal or corrected-to-normal vision, and normal or near-normal hearing. Vision was tested by identifying the visual 'catch stimulus' in a 3 × 3 grid where the other eight stimuli were in the 'normal' condition. The visual stimuli used for this purpose were the same as the stimuli used during data collection. Participants were seated at a viewing distance of 1.5 m and they had to do this task three times correctly before being

allowed to participate. There was no time limit during this vision test and the visual stimulus was repeatedly played until the odd stimulus was identified.

The inclusion criterion for normal hearing was having hearing thresholds lower than or equal to 25 dB HL at the audiometric test frequencies of 0.25, 0.5, 1, 2, and 4 kHz for both ears, measured with standard clinical audiometry procedures. For the young group, 21 individuals (4 males), all below the age of 30 years (23.4 yr $\pm$ 3.2), participated in the study. For the older group, 64 older individuals with self-reported normal hearing were screened. Among the older individuals, 47, who initially volunteered to participate, did not meet the inclusion criterion for normal hearing, and were therefore excluded from the study before testing. From the remaining 17 older participants, two were not able to do the task, leading to their exclusion during the testing. After the exclusions, the older group consisted of 15 participants (3 males), all above the age of 60 years (64.5 yr $\pm$ 2.7).

Figure 1 shows the average hearing thresholds for the two groups (Y = young; O = older). Despite the careful hearing screening, there was a small difference in the thresholds between the two groups (which we have also observed in our previous studies on age effects, e.g., Saija *et al.*, 2014, 2019). As a precaution, to explore potential audibility effects, we investigated this difference. Firstly, we focused on the audiometric test frequency of 500 Hz, the frequency of the pure tone stimulus used in the study. At this test frequency, the average hearing thresholds were 2.3 dB $\pm$ 3.3 and 7.2 dB $\pm$ 6.2 for young
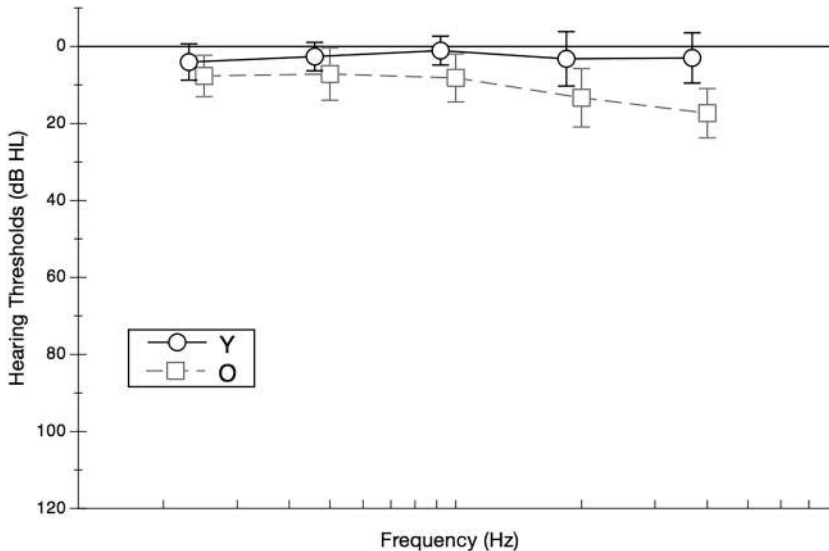
Figure 1. Hearing thresholds shown for the young (Y) and older (O) groups, averaged over the participants and the two ears.

and older groups, respectively, which did not differ significantly ($p = 0.072$, by a Mann–Whitney U test). Secondly, we focused on the audiometric test frequencies between 0.25 and 4 kHz, as this range corresponded to the bandwidth of the lowpass-filtered speech stimuli used in the study. At these test frequencies, the average hearing thresholds were 2.3 dB $\pm$ 3.0 and 10.7 dB $\pm$ 4.7 for young and older groups, respectively. While the hearing thresholds differed significantly between young and older individuals ($t = -5.685$, $p < 0.001$, by a two-tailed t-test with unequal variances), none of the older participants showed hearing threshold deficits larger than 20 dB, rendering all older participants as hearing within normal limits. The average interaural threshold differences were almost identical, 4.5 dB $\pm$ 1.3 and 4.6 dB $\pm$ 1.7 for young and older groups, respectively.

The Medical Ethical Committee of the University Medical Center Groningen approved the study protocol. Before the screening, the participants received written and oral information about the study and provided written informed consent. They were reimbursed for travel expenses and participation time according to departmental policy.

2.2. Auditory Stimuli

Two types of auditory stimuli were used, pure tone and speech. The pure-tone stimulus consisted of four 200-ms long 500-Hz tone bursts with an interval of 1 s between the individual bursts. Each tone burst had a 5-ms on/off ramping by Hann window. The pure-tone stimulus was presented at the sensation level (SL) of 60 dB re individual hearing threshold measured at 500 Hz and averaged over both ears. By presenting the stimuli at the same SL across participants we aimed to account for the listener-specific hearing thresholds, which slightly differed across participants (as described in the previous section). The speech stimulus consisted of digital recordings of meaningful consonant–vowel–consonant (CVC) Dutch words, spoken by a female speaker, and taken from the corpus of the Nederlandse Vereniging of Audiologie (NVA; Bosman and Smoorenburg, 1995). We chose this corpus as it is also used as a clinical diagnostic tool with hearing-impaired populations in Dutch clinics. The corpus has 180 unique words that are ordered into 15 unique lists of 12 words. In clinical assessments, the number of lists is usually increased to 45 by re-ordering the words within a list, and such an extended corpus was also used in our study. The lists are balanced across each other in phonemic distribution. The duration of words ranges roughly between 700 and 1000 ms. The speech materials used in our study were lowpass-filtered (3-kHz cutoff frequency, 60-dB/octave slope) to further ensure similar audibility between young and older groups, as hearing thresholds at audiometric test frequencies above 4 kHz were not part of the inclusion criteria. Similar to tones, the speech stimuli were presented at

the individually adjusted SL of 60 dB re average individual hearing thresholds at 0.5, 1, and 2 kHz for both ears.

2.3. Binaural Stimulus Reproduction

Acoustic targets were created by filtering auditory stimuli (in case of speech stimulus, following the low-pass filtering) with spatially up-sampled HRTFs of the KEMAR manikin (Gardner and Martin, 1995). Listener-specific HRTFs were not required because the spatial direction of the virtual stimuli varied only along the horizontal plane and non-individualized HRTFs are thought to provide sufficient cues for the sound localization in horizontal planes (Wenzel *et al.*, 1993).

The original HRTFs from the KEMAR manikin were available (Note 2) at a lateral resolution of 5° (see Fig. 2, top panel). This lateral sampling is larger than the MAAs found in normal-hearing listeners (approximately 1° — Perrott and Saberi, 1990). Thus, the original HRTFs were not sufficient for our study. A super-resolution HRTF set was calculated by directionally up-sampling the original HRTF set to the lateral resolution of 0.5°. The most salient cues for the lateral direction of a sound are the broadband interaural time and level differences (ITDs and ILDs, respectively; Macpherson and Middlebrooks, 2002). Correspondingly, for each ear, in the original HRTF set, the broadband timing and amplitude spectra were directionally interpolated. More specifically, for each ear's HRTF set, broadband timing was removed, amplitude spectra were interpolated, and the interpolated timing information was applied. The broadband timing was removed by replacing the HRTF's phase spectrum by the minimum-phase spectrum (Oppenheim *et al.*, 1999) corresponding to HRTF's amplitude spectrum. For the interpolation of the amplitude spectra, the complex spectra of the minimum-phase HRTFs for two adjacent available directions were averaged according to a weighting that corresponded to the interpolated target direction.

For the interpolation of the timing, a continuous-direction model of the time-of-arrival (TOA) was applied (Ziegelwanger and Majdak, 2014). TOA is the broadband delay arising from the propagation paths from the sound source to the listener's ear. For a given direction of a sound, the interaural difference of TOAs corresponds to the ITD. The TOA model parameters describe listener's geometry (head and ears) and configure a continuous-direction function of broadband TOA. We used this function to calculate TOAs for directions in steps of 0.5°. To this end, for each ear, the model was fit to an HRTF set as described by Ziegelwanger and Majdak (2014) using the implementation from the Auditory Modeling Toolbox (Søndergaard and Majdak, 2013). Then, each minimum-phase HRTF was temporally up-sampled by a factor of 64, circularly shifted by the TOA obtained from the continuous-direction TOA model for the target direction, and then down-sampled to the sampling rate of
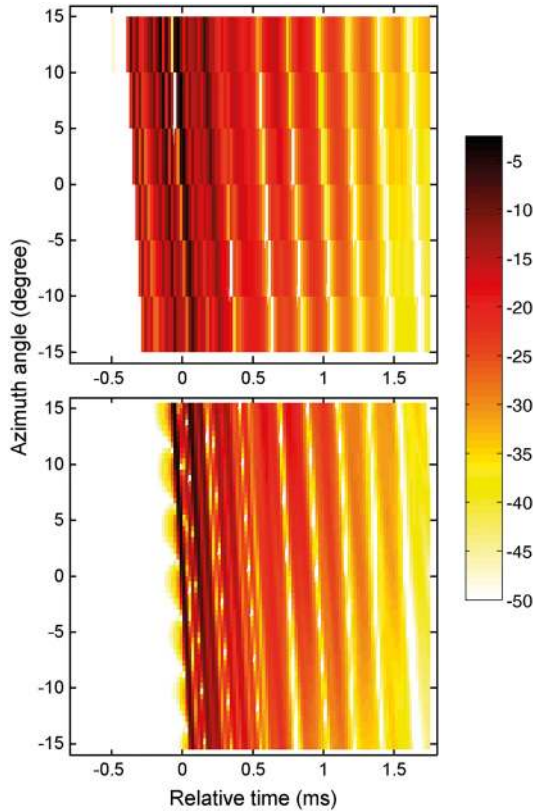
Figure 2. Left-ear head-related transfer functions (HRTFs) shown in the time domain (i.e., head-related impulse responses) as a function of the azimuth angle. Top: Original HRTFs (resolution of 5°; Gardner and Martin, 1995). Bottom: Interpolated HRTFs (super resolution of 0.5°). Color: Amplitude of the impulse responses shown in dB.

44.1 kHz (Fig. 2, lower panel). Note that the temporal oversampling was required to achieve an interaural resolution of 0.35 μs. A brief quality check (see Fig. 2), revealed (1) the main peaks at the same temporal positions as those in the original HRTFs, and (2) similar temporal modulations in both original and super-resolution HRTFs. Note that, as a result of the conversion to minimum-phase systems, the slowly rising energy before the main peak present in the original HRTFs is not present in the super-resolution HRTFs. In summary, the final HRTF set (Note 3) contained HRTFs with the interpolated amplitude and broadband timing information associated with the ILD and broadband ITD, respectively, at a lateral resolution of 0.5°.
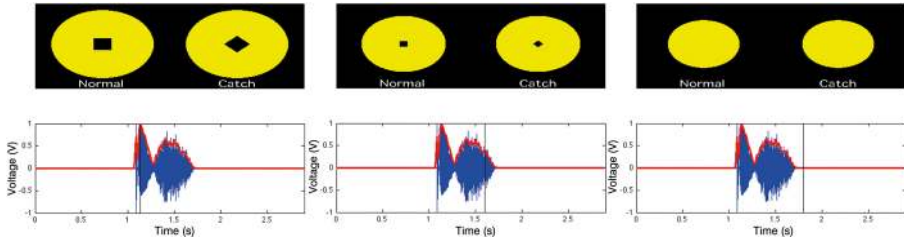
Figure 3. Snapshots of the visual stimulus (top row) shown with the corresponding auditory speech stimulus (CVC word 'poes'; bottom row). Top: Visual stimuli in normal and catch trials are shown, alternating in each panel from left to right, and each panel shows a different snapshot taken at a different point in time. Bottom: The auditory stimulus is shown in temporal waveform. The red contour line shows the slow-moving envelope of the auditory signal over time. Black vertical lines mark the point in time of the snapshots. Note the correspondence between the square size of the visual stimulus and the envelope amplitude in the auditory stimulus at the specific times shown with the vertical black lines.

2.4. Visual Stimulus

The visual stimulus was the same geometric shape for both tone and speech stimuli. This shape was modulated according to auditory signal intensity, which differed between the tone and speech stimuli. We opted to use the same simple visual stimulus for both stimulus types, instead of using lipreading cues for speech, for several reasons: (1) to ensure consistency between the two stimulus types, (2) to ensure simplicity, for example for potential clinical applications, where it would be easier to implement a generic visual stimulus, and (3) to minimize any potential interference from additional cognitive processing that may be required from speech lipreading. The generic visual stimulus consisted of a yellow circle on a black background presented in the center of the screen (Fig. 3). The diameter of the circle was modulated in proportion to a 16-ms moving average of the root-mean-square (RMS) amplitude of the auditory stimuli, with a minimum size of 10 mm and a maximum size of 15 mm. Further, a black square was shown on top of the yellow circle, in the center. The edge length of the square was proportional to the RMS amplitude of the auditory signal, with a minimum size of 0 mm and a maximum size of 3 mm. The size of the objects followed the sound amplitude immediately, being only limited by the update rate of the computer monitor. In the catch trials (explained later), the square was rotated by 45°. In order to focus attention on the screen, visual rendering started 1 s prior to the auditory stimulus and showed the yellow circle with minimum size until the auditory stimulus had started.

2.5. Apparatus

The experiment was conducted in an anechoic chamber. Participants were seated in a chair located at a distance of 1 m from the computer screen. The

chair was specifically designed for this study. An adjustable neck rest limited head movement, and ensured that the participant faced the screen that displayed the visual stimulus.

The auditory stimuli were digitally created with a sampling rate of 44.1 kHz using Matlab R2009b (Mathworks Inc., Natick, MA, USA) on a Mac computer (Apple Inc.). They were routed via the digital sound interface Audiofire 4 (Echo Digital Audio Corporation, Santa Barbara, CA, USA) to the digital-to-analog converter DA10 (Lavry Engineering Inc., Kingston, WA, USA) and then presented via the headphones HD 600 (Sennheiser, Wedemark, Germany). The presentation levels of the auditory stimuli were calibrated with a KEMAR manikin (GRAS, Holte, Denmark) and the sound level meter Type 2610 (Brüel Kjær, Nærum, Denmark). The linearity of the system was verified for SPLs between 40 and 90 dB, i.e., the SPL range of our auditory stimuli. The visual stimulus was presented on a computer screen, with an update rate of 60 frames per second. For presentation of auditory and visual stimuli, PsychToolBox-3 (Kleiner *et al.*, 2007) was used. This software is designed in particular to allow a synchronized auditory and visual presentation, with an intermodal timing accuracy of approximately 2 ms. We have confirmed this with multiple measurements. Audio stimulus delay was quantified by time-stamping the command to present a signal and time-stamping incoming audio on a microphone, mounted on the KEMAR manikin. The delay measured was less than 1 ms. Video stimulus delay was quantified by internal diagnostics by time-stamping the command to display a target and obtaining the timestamp of the completed visual rendering. This delay was also less than 1 ms. Even when multiple visualization commands and audio commands were issued during a testing sequence, a delay of more than 1 ms was never measured. Thus an intermodal timing accuracy of less than 2 ms was obtained. We have not conducted any other controls for other potential delays. We used PsychToolBox-3 to create the intermodal lags that were part of the experimental design.

2.6. Procedure

All participants were naïve to the experimental protocol. All tests were administered and evaluated by the first author.

MAAs in the horizontal plane were measured using a lateralization task in an adaptive 3-down-1-up staircase procedure (Levitt, 1971); however, two runs, starting both at extreme lateral positions for left and right, were simultaneously interleaved, following the procedures of Bertelson and Aschersleben (1998). In each trial, a target auditory stimulus was presented, with or without visual stimulus, depending on the experimental condition. Participants were asked to make a left/right judgment according to where they lateralized the target source by saying 'links' or 'rechts' (Dutch equivalent of 'left' and 'right', respectively). Each run started with an auditory target virtually positioned at

the lateral angle of 10°. After three consecutive correct responses, the angle decreased by a fixed initial step size. After an incorrect response, the angle increased. The transition from decreasing to increasing angle, and vice versa, defined a reversal. The initial step size was 4°, and with each reversal, the step size was halved until the minimum of 0.5°, the resolution of our modified HRTFs, was reached. The trials from the two interleaved runs were randomly chosen such that the participant was not aware of the side actually tested. Both runs continued until eight reversals were obtained for each. For both runs, the values measured at the last four reversals were averaged and the difference between the averages produced the MAA. Depending on testing conditions and participant performance, an MAA was acquired in 5 to 15 minutes, after which a pause was given.

2.7. Testing Conditions

The MAA was measured for each auditory stimulus (pure tone, speech) in combination with three AV conditions, namely, auditory only (NoV), with synchronous visual stimulus (SyncV), and with asynchronous visual stimulus (AsyncV).

Participants were asked to look straight ahead to the monitor while performing the lateralization task, and movement of the head was prevented with the head rest on the chair. During the NoV condition, only the auditory signal was presented, with no visual stimulus. Participants then saw the monitor that was turned off or had the option of having their eyes closed. In AV conditions, the stimulus was played on the monitor placed at 0°. Catch trials were used to make sure in AV conditions participants did not have their eyes closed. During the SyncV condition, the auditory signal was presented synchronously with visual stimulus. During the AsyncV condition, the auditory signal randomly lagged or led with respect to the onset time of the visual stimulus. The lag/lead duration was in the range of 400 to 500 ms, which provides a noticeable asynchrony (Alm and Behne, 2013; Başkent and Bazo, 2011; Hay-McCutcheon *et al.*, 2009). To make sure that attention was given to the visual stimulus during both AV conditions of SyncV and AsyncV, catch trials were introduced at an occurrence chance level of 20% of the trials (Fig. 3). In the catch trials, the participants had to identify the change in the orientation of the black square in the visual stimulus by saying 'ja' ('yes'). If the participant failed to identify catch trials two consecutive times or failed to identify them more than twice in total, the run was declared invalid and was repeated until a successful completion. Using these catch trials, we identified the two older participants who were not able to do the task and they were consequently excluded from the experiment.

All six conditions ([pure tone, speech] × [NoV, SyncV, AsyncV]) were tested in one day, and for each participant, three MAAs were recorded per

condition, each on a separate day. The order of the six conditions was determined by a normalized Latin-square design. In the case of speech stimuli, the list order and the word order in each list were randomized. The total testing time per day was approximately two hours.

3. Results

3.1. Task Progress

Figure 4 shows an example of two interleaved runs for a participant from the older group. Note the successive decrease in lateral position of the target converging at the threshold in both runs. For each run, a threshold is denoted by the corresponding horizontal dotted line, and the MAA is the difference between the two thresholds. While this exemplary participant seems to show a left bias in this run, informal checks of other participants did not reveal a systematic bias.

3.2. Baseline Comparison for the Age Groups

Figures 5 and 6 show the MAA statistics for the pure-tone and speech stimuli, respectively, for the two age groups tested under the three AV conditions (NoV, SyncV, AsyncV). First, we investigated the baseline auditory-only MAA measurements between Y and O groups, to confirm these were comparable by analyzing the NoV results only. For the reliability of our task, we first analyzed the MAA standard deviations (SDs) in the NoV conditions. For pure-tone
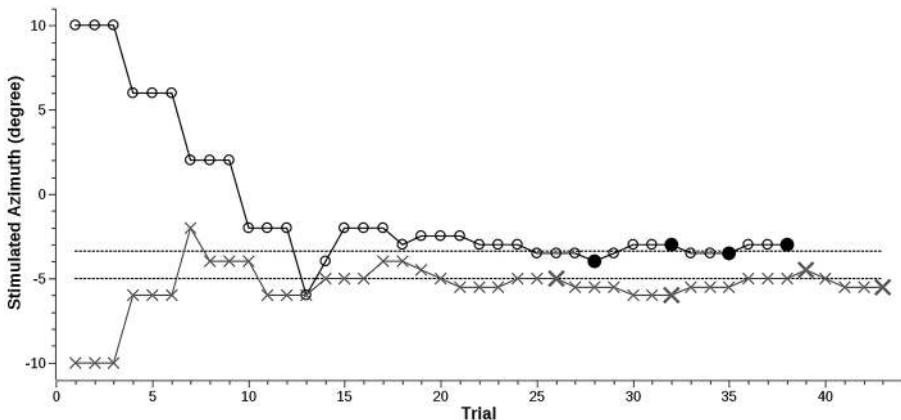
Figure 4. An example of interleaved runs. The black line with circles and the grey line with crosses indicate right- and left-initiated runs, respectively. The large filled circles and large bold crosses indicate reversals used for averaging for each run. The resulting thresholds are shown by the dotted lines for each run. Minimum audible angle (MAA) was calculated from the difference between these two thresholds. This participant shows a left bias in this specific interleaved run.
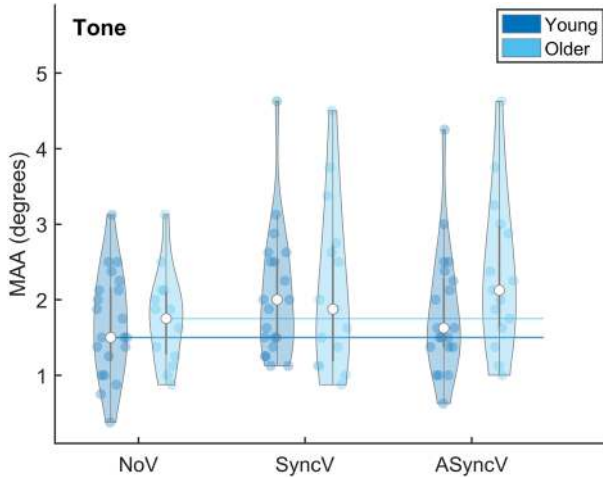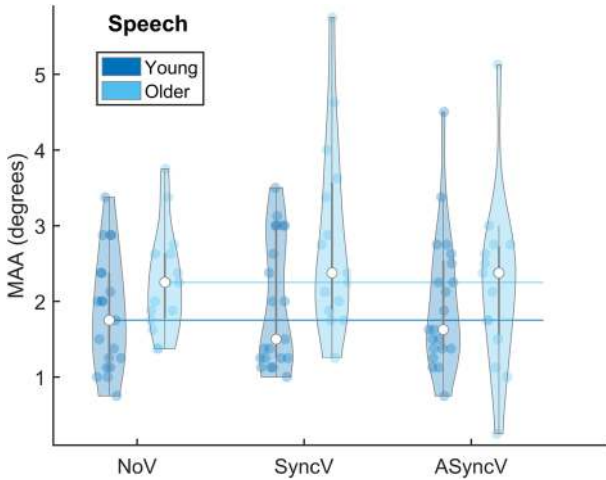
Figure 5. MAAs shown for the three audiovisual (AV) conditions in response to pure-tone stimulus. From left to right are shown the MAAs with auditory-only baseline condition with no visual stimulus (NoV), with visual stimulus presented synchronously with auditory stimulus (SyncV), and with visual stimulus presented asynchronously with auditory stimulus (AsyncV). For each AV condition, MAAs from young and older individuals are shown together on the left and right, respectively. For each participant group and each AV condition, the open circles show the average and the filled circles show the individual data. The shape follows the kernel density plot estimated for the data. The thick vertical lines show the interquartile range. The horizontal lines show the baseline auditory-only NoV average MAAs as references for easier comparison to conditions with the visual stimulus added (SyncV and AsyncV).

stimulus, the SDs were $0.69° \pm 0.23°$ for younger and $0.63° \pm 0.25°$ for older groups, with no significant difference [$F(14, 20) = 0.75$, $p = 0.60$], and for speech stimulus, $0.70° \pm 0.23°$ for younger and $0.73° \pm 0.20°$ for older groups, also with no significant difference [$F(14, 20) = -0.86$, $p = 0.78$]. With the similar variance between the groups, we analyzed the MAAs with an unpaired *t*-test for equal variance. For the pure-tone stimulus, group-averaged MAAs were $1.67° \pm 0.69°$ for young and $1.76° \pm 0.59°$ for older groups, with no significant between-group difference [$t(34) = 0.39$, $p = 0.70$]. For the speech stimulus, group-averaged MAAs were $1.85° \pm 0.74°$ for young and $2.26° \pm 0.69°$ for older groups, also with no significant difference [$t(34) = 1.70$, $p = 0.10$]. Overall, this analysis confirms comparable baseline performances between the two groups, indicating the suitability of our paradigm to test young and older participants.

3.3. General Overview

For a general overview of results, we analyzed all MAA measurements together using a three-way mixed-model ANOVA, with as within-subject factors stimulus type (pure tone or speech) and AV condition (NoV, SyncV, AsyncV),

Figure 6. MAAs shown for the three audiovisual (AV) conditions in response to speech stimulus. From left to right are shown the MAAs with auditory-only baseline condition with no visual stimulus (NoV), with visual stimulus presented synchronously with auditory stimulus (SyncV), and with visual stimulus presented asynchronously with auditory stimulus (AsyncV). For each AV condition, MAAs from young and older individuals are shown together on the left and right, respectively. For each participant group and each AV condition, the open circles show the average and the filled circles show the individual data. The shape follows the kernel density plot estimated for the data. The thick vertical lines show the interquartile range. The horizontal lines show the baseline auditory-only NoV average MAAs as references for easier comparison to conditions with the visual stimulus added (SyncV and AsyncV).

and as between-subject factor, group (young, older). There was a significant main effect of stimulus type [$F(1, 34) = 4.80$, $p = 0.035$, $\eta^2 = 0.0027$] and AV condition [$F(2, 68) = 5.96$, $p = 0.004$, $\eta^2 = 0.0051$], and a significant three-way interaction [stimulus type × AV condition × group, $F(2, 68) = 3.36$, $p = 0.041$, $\eta^2 = 0.0029$].

The significant main effect of stimulus type and the significant interaction of the main effects led us to perform the following investigations per stimulus type.

3.4. Results for Pure-Tone Stimuli

Figure 5 shows the group-averaged MAAs for pure-tone stimuli, for the two participant groups and for all three AV conditions. A two-way mixed-model ANOVA was conducted with the between-subject factor of group (young, older) and the within-subject factor of AV condition (NoV, SyncV, and AsyncV). There was a significant main effect of AV condition [$F(2, 68) = 4.63$, $p = 0.013$, $\eta^2 = 0.0079$], but no significant main effect of group [$F(1, 34) = 0.92$, $p = 0.34$, $\eta^2 = 0.0033$] and no significant interaction [$F(2, 68) = 1.62$, $p = 0.21$, $\eta^2 = 0.0028$]. A multiple comparison test (based

on the Tukey–Kramer honestly significant difference procedure) showed that both SyncV and AsyncV conditions yielded significantly ($p < 0.035$) larger MAAs than the NoV condition, indicating a significant ventriloquist illusion for both. All other compared pairs of conditions were not significantly different ($p > 0.05$).

3.5. Results for Speech Stimuli

Figure 6 shows the group-averaged MAAs for speech stimuli, for the two participant groups and for all three AV conditions. A two-way mixed-model ANOVA was conducted with the between-subject factor of group (young, older) and the within-subject factor of AV condition (NoV, SyncV, and AsyncV). There was no significant main effect of group [$F(1, 34) = 3.83$, $p = 0.059$, $\eta^2 = 0.013$] or AV condition [$F(2, 68) = 2.88$, $p = 0.063$, $\eta^2 = 0.0049$], and no significant interaction [$F(2, 68) = 2.22$, $p = 0.12$, $\eta^2 = 0.0038$].

4. Discussion

Our results show that the ventriloquist illusion can be elicited with virtual spatial cues by using generic and easily available HRTFs. Even though the current experiment used an anechoic chamber, to provide clean baseline data for future studies, our application of HRTFs indicates the illusion can be used without a need for elaborate multi-speaker setup and an anechoic chamber. Our paradigm, based on MAA measurements with interleaved adaptive procedures, showed a ventriloquist illusion with a pure-tone stimulus, not only with a synchronously presented but also with an asynchronously presented visual stimulus. On average, the illusion was observed in a similar manner in both young and older listeners, with individuals selected to minimize age-related hearing loss effects, and stimuli adjusted to minimize potential audibility effects. Further, the baseline auditory-only performance did not significantly differ between the young and older groups. All observations combined suggest that the modified ventriloquist illusion paradigm may be a useful tool that is simple to implement and easy to use to systematically investigate AV integration in young and older individuals.

4.1. Virtual Spatial Cues: Binaural Stimulus Production and HRTFs

The MAAs measured with binaural stimulus reproduction via HRTFs were in the range of a few degrees, in line with the previously reported human ability to discriminate spatial cues in the horizontal plane (Middlebrooks and Green, 1991). Adding the visual stimulus resulted in a small (around half a degree on average) but significant increase in measured MAAs. These results indicate that the ventriloquist illusion can be elicited using virtual spatial cues

via headphones. Hence, it seems that the AV binding leading to the illusion can occur even when listening to *non-individualized* generic HRTFs, indicating that an externalized (out-of-the-head) perception of the virtual sound is not required. Note that our task was restricted to the horizontal plane where broadband interaural cues are thought to be the most salient (Macpherson and Middlebrooks, 2002). Thus, our findings would not necessarily apply to a ventriloquism task performed in vertical planes, where listener-specific spectral cues would be important for sound localization and externalization (Langendijk and Bronkhorst, 2002). However, for inducing ventriloquist illusion in the horizontal plane, the easily available generic HRTFs seems a useful tool, helping with a simpler implementation of the ventriloquist paradigm.

4.2. *Stimulus Type: Pure Tone* versus *Speech*

We have explored the ventriloquist illusion with both pure-tone and speech auditory stimuli. For both auditory stimulus types, for consistency and simplicity, we have used the same visual stimulus type, which was a geometric shape modulated in accordance with the auditory stimulus presentation level. Despite the consistency of using the same visual geometric shape, the illusion was observed only for a pure-tone stimulus, and not for speech. Pure-tone stimuli are more simplistic in nature, likely inducing more of the automatic processes. Speech, on the other hand, is a complex signal, more ecologically valid, and highly learned due to exposure from daily communication. Only under ideal conditions (no background noise, no hearing disorder, clear pronunciation of speech by a native speaker, etc.) is perception of speech considered an automatic process — for any deviation from ideal listening it likely requires more cognitive processes (Mattys *et al.*, 2012; Wild *et al.*, 2012). Previous research on other forms of AV integration or illusion tasks indeed also showed differential effects with stimuli of varying complexity (e.g., Vatakis and Spence, 2006) or between speech and non-speech stimuli (e.g., Tremblay *et al.*, 2007), and this was used as a partial explanation for inconsistent reports of age on AV integration (e.g., Laurienti *et al.*, 2006; Stevenson *et al.*, 2015; Tye-Murray *et al.*, 2010).

Our results with the two auditory stimulus types were in line with these ideas. Our statistical analyses showed not only a significant effect of adding the visual stimulus, but also a significant effect of the stimulus type. Therefore, the results were re-analyzed separately for pure-tone and speech stimuli, and these analyses indicated a ventriloquist illusion with tone stimuli, but not with speech stimuli. The observed illusion with the simpler auditory stimulus of pure tone is in line with the idea that the ventriloquist illusion is mainly pre-attentive and relies on automatic processes (e.g., Bertelson *et al.*, 2000; Vroomen and de Gelder, 2004), and the illusion with speech is perhaps more

attention-related (e.g., Driver, 1996). On the other hand, the well-known real-life manifestation of the ventriloquist illusion is where a listener is convinced that a puppet with a synchronously moving mouth piece is talking, hence, we know that the illusion works for speech and in ecologically valid settings. The lack of illusion for speech stimuli in our study, hence, could be due to factors related to our experimental design. One difference between the tone and speech stimuli was their total duration. While the tone stimulus was short, 200 ms, it was repeated four times, with relatively long inter-tone duration of 1 s, producing tone burst sequences of 3800 ms. Speech stimuli, in contrast, varied roughly between 700 and 1000 ms in duration, much shorter than the tone sequences. While this choice was the result of aiming for simplicity, as well as using a clinically relevant material (where the speech stimuli were taken from a typical clinical speech test), it is possible the speech recordings were too short in duration to induce the illusion. Perhaps with longer stimuli, such as sentences, there would be a build-up to illusion. Further, the choice for using a geometric shape as visual stimulus for speech, driven by consistency and simplicity purposes, might have affected the results. Since the pure-tone stimulus, once it was on, did not change in its intensity level, the accompanying visual stimulus was rather static during the on times, and the biggest changes occurred at tone burst onset and offset. The human auditory system is sensitive to such onsets and offsets, and it is possible that this combination was useful in inducing the illusion. For speech, the movements of the geometric shape were more dynamic as the intensity of the signal varied not only at word onsets and offsets, but also during the utterance. We had assumed that such dynamic features would help with stronger AV binding, but our data did not support this expectation. It is possible that actual face or mouth movements would be better visual stimuli for inducing the illusion with speech, as would be the case with puppets. For example, Driver (1996) showed strong illusion effects with speech stimuli that were longer than our stimuli (three words *versus* one word), when presented with actual lipreading cues from full face recordings as visual stimuli. On the other hand, studies on the McGurk effect, another AV illusion that heavily relies on speech phoneme perception but likely on different perceptual/neural AV integration mechanisms (McGurk and MacDonald, 1976; for a review, see Alsius *et al.*, 2018), showed that a full visual representation of the face is not required. For example, Rosenblum and Saldaña (1996) showed the McGurk illusion even when the face was represented by a point-light display. More recent studies, such as Files *et al.* (2015), indicated that visual speech is represented by both its motion and configurable attributes, results found from using synthetic visual speech stimuli. Hence, given that these examples of AV integration may rely on other mechanisms than those responsible for our ventriloquist illusion, it remains

unclear if a more face-like or lip movements-like visual stimulus would have induced stronger ventriloquist illusion than a face-unlike geometrical shape.

Overall, our results readily support the idea that a tone auditory stimulus and a geometric visual stimulus can be used for a relatively simple implementation of ventriloquist illusion to explore AV integration; however, more fine-tuning is needed to investigate use of speech materials for this purpose.

4.3. *Synchrony* versus *Asynchrony of the Visual Stimulus*

We have tested the ventriloquism effect with synchronous and asynchronous A and V stimuli in order to investigate whether temporal synchronicity modulates the AV integration differently for younger and older individuals (e.g., Alm and Behne, 2013; Başkent and Bazo, 2011; Diederich *et al*., 2008; Hay-McCutcheon, 2009). Our statistical analyses did not show any evidence for the effect of synchronicity. In fact, for pure-tone stimuli, both visual conditions with synchronous and asynchronous presentation yielded significantly larger MAAs than those obtained in the auditory-only baseline condition.

The lack of effect of synchronicity is an interesting observation because our asynchrony (ranging between 400 and 500 ms) was larger than the asynchrony thresholds of a few hundred milliseconds previously reported (Alm and Behne, 2013; Başkent and Bazo, 2011; Grant and Seitz, 1998; Hay-McCutcheon, 2009; Massaro *et al*., 1996). There might be several explanations for this discrepancy. In those studies, the stimuli used were mostly speech, and the task for the participant was to report the point of synchrony/asynchrony distinction for audiovisual speech presented from one location. In contrast, our ventriloquist effect appeared with pure tones, and the task of our listeners was to report the perceived location of a sound source, with or without the accompanying visual stimulus. The temporal AV integration window is most likely dependent on the specific stimuli and task used (e.g., Stevenson and Wallace, 2013), and perhaps for the illusion the integration window was longer. Regardless, the lack of an effect from the very long asynchrony introduced between the auditory and visual stimuli on the ventriloquist illusion indicate once more the robustness of the effect. For practical implications, a test based on this illusion would then not be expected to be negatively affected by a potential asynchrony that may be caused by software or hardware settings and limitations.

4.4. *Age Effects*

In auditory-only baseline conditions with no visual stimulation (NoV), our older participants performed similarly to the young participants, in agreement with studies showing only minor age-related changes in sound localization in the horizontal plane (Abel *et al*., 2000; Otte *et al*., 2013). Having comparable auditory-only baseline performance between young and older groups allows a

fair comparison of changes due to addition of visual cues, reducing the confound of inverse effectiveness. This way, by utilizing the MAAs, ventriloquist illusion presents a potentially useful tool to investigate AV integration.

Our statistical analyses did not provide evidence for a significant group difference. Both groups showed a significant increase in MAAs with the addition of both synchronously and asynchronously presented visual stimuli for tone stimulus (which we took as the evidence for the ventriloquist illusion), and a non-significant increase in MAAs for speech stimulus (which we took as the lack of illusion). Hence, while these findings supported the idea that the ventriloquist illusion can be induced with MAAs implemented using HRTFs in both young and older individuals, the results *per se* did not indicate an age effect on AV integration.

Previous studies on aging and AV integration indicated differing motivations for why age could have an effect on AV integration. One idea has been that older individuals may show greater gain of multimodal stimuli compared to unimodal stimuli as a result of compensation for age-related sensory and cognitive changes that would affect perception in general (e.g., Laurienti *et al.*, 2006). Others have argued that, as a result of age-related inhibition the effect from a stimulus presented from another modality may have a larger effect on perception in older adults than young adults (e.g., Couth *et al.*, 2018). An increased temporal integration was also suggested, as a result of a general age-related slowing down (e.g., Pfeiffer *et al.*, 2007), which may lead to stronger AV binding of sequentially presented multimodal stimuli (Alm and Behne, 2013; Başkent and Bazo, 2011; Hay-McCutcheon *et al.*, 2009). Some studies indeed indicated a stronger AV integration in older individuals, but these often were confounded by inverse effectiveness [e.g., when measured in response times (Laurienti *et al.*, 2006); when measured in speech intelligibility, and also in the presence of age-related hearing loss (Başkent and Bazo, 2011)]. In contrast, some studies showed a smaller benefit from AV integration in older adults (e.g., with degraded visual stimulus quality; Tye-Murray *et al.*, 2010), but sometimes it was not possible to tease apart the effect of aging from the effect of age-related hearing loss (e.g., when measured in speech intelligibility; Musacchia *et al.*, 2009).

In our study, our participants were selected to have almost no hearing loss and corrected vision, and further, only individuals who could do the experimental task participated. Hence, no or minimal effects were expected from age-related sensory or cognitive changes. Further, the baseline performance with no visual stimuli was the same between young and older groups, and the task did not depend on speech intelligibility, for which age-related deficits in lip-reading may play a role (e.g., Cienkowski and Carney, 2002; Sommers *et al.*, 2005). One reason for the lack of an age effect, as different from what is described in the literature, could be that we have controlled for all the other

potential factors than age that can lead to an effect on AV integration. Another reason might be that the use of the ventriloquist illusion paradigm, which potentially relies on automatic processes, may be less sensitive to age-related changes in cognitive mechanisms.

4.5. Clinical Relevance

Aging is often accompanied with age-related changes in sensory (e.g., hearing impairment) and cognitive capabilities (e.g., working memory, processing speed), both of which can affect mechanisms of multisensory integration. Multisensory integration is considered to be closely linked to the ability to conduct activities of daily living, especially for older individuals (de Dieuleveult *et al*., 2017; Basharat *et al*., 2018; and see, for example, for balance and falling, Mahoney *et al*., 2014; Setti *et al*., 2011). Therefore, the search for practical, applicable, and effective tests for multisensory integration, which can be implemented in clinical settings, continues (e.g., de Dieuleveult *et al*., 2019).

The present study concerns a specific form of multisensory integration, namely audiovisual integration, which can be affected by age-related hearing loss. Clinical assessment of hearing impairment typically involves pure-tone audiometry, which relies on measuring hearing thresholds of tones presented at differing center frequencies, or speech audiometry, which often relies on hearing and understanding simple phonemes or words (e.g., Katz, 2014). The former is used in defining the degree and type of hearing loss, while the latter is used as an indication for the functional effect of hearing impairment on speech communication. Yet, daily speech communication rarely occurs in the auditory domain only. In fact, it often involves the integration of visual cues into speech perception in order to enhance the overall intelligibility performance, especially in hearing-impaired individuals (e.g., Erber, 1975). Hearing-impaired individuals present a wide range of inter-individual cognitive compensation (e.g., Başkent *et al*., 2016) and AV integration skills (Altieri and Hudock, 2014; Başkent and Bazo, 2011). Such variation likely results in varying degrees of success in enhancing the auditory speech performance by using visual cues. Still, clinical tests are not capable to capture such individual integration variability yet.

For a more comprehensive assessment of real-life communication performance, as well as other daily activities that may depend on AV integration, one would ideally like to add a simple test of AV perception. Freiherr *et al*. (2013) and de Dieuleveult *et al*. (2019), for example, argue for the importance of clinical tests that can identify changes in multisensory integration sufficiently early, such that the best individualized therapies and support tools can be offered to older individuals or patients. Yet, such attempts can be hindered by obstacles such as convoluted tasks for these individuals, complex

measurement setup, and the limited time a clinician can spend with each patient. Therefore, for a realistic transfer of a new test into the clinical domain, the test needs to be easy to administer, and be able to produce reliable results within a reasonable duration of time.

The method proposed in this study has such a potential for future clinical applications. The advantage of our method is that it is an easy task, independent of speech intelligibility and potentially less sensitive to cognitive processes. The setup is simple, requiring only headphones and a set of publicly available HRTFs (see link in Note 3). While in its current form the testing time was not yet very short, one should note that this is mainly caused by use of two sets of stimuli, and a large number of reversals and repetitions, which could all potentially be optimized. In order to explore potential clinical applicability and to reduce the overall test time while maintaining test reliability, all of these factors need to be critically evaluated and optimized for target groups of interest in follow-up studies.

Notes

1. Multisensory integration is the process where information from multiple senses is combined to produce a single coherent percept. This process, however, can include a number of mechanisms, such as statistical facilitation and vigilance, in addition to a core neural integration of multisensory data (e.g., Colonius and Arndt, 2001; Van Opstal, 2016). In the present study, we use the term multisensory integration in a broader sense than and relatively independent from the specific underlying neural mechanisms, focusing on the effects observed on one sensory modality (auditory) when presented together (in temporal overlap or proximity) with another sensory modality (visual), and as observed in global behavioral data (e.g., Chen and Vroomen, 2013).

2. Available at http://sound.media.mit.edu/resources/KEMAR.html, last retrieved on February 03, 2019.

3. The original and the interpolated HRTF sets are both available as SOFA files (Majdak et al., 2013) at https://doi.org/10.5281/zenodo.3250072.

References

Abel, S. M., Giguère, C., Consoli, A. and Papsin, B. C. (2000). The effect of aging on horizontal plane sound localization, *J. Acoust. Soc. Am.* **108**, 743–752.

Alais, D. and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration, *Curr. Biol.* **3**, 257–262.

Alm, M. and Behne, D. (2013). Audio-visual speech experience with age influences perceived audio-visual asynchrony in speech, *J. Acoust. Soc. Am.* **134**, 3001–3010.

Alsius, A., Paré, M. and Munhall, K. G. (2018). Forty years after hearing lips and seeing voices: the McGurk effect revisited, *Multisens. Res.* **31**, 111–144.

Altieri, N. and Hudock, D. (2014). Assessing variability in audiovisual speech integration skills using capacity and accuracy measures, *Int. J. Audiol.* **53**, 710–718.

Basharat, A., Adams, M. S., Staines, W. R. and Barnett-Cowan, M. (2018). Simultaneity and temporal order judgments are coded differently and change with age: an event-related potential study, *Front. Integr. Neurosci.* **12**, 15. DOI:10.3389/fnint.2018.00015.

Başkent, D. and Bazo, D. (2011). Audiovisual asynchrony detection and speech intelligibility in noise with moderate to severe sensorineural hearing impairment, *Ear Hear.* **32**, 582–592.

Başkent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., Sarampalis, A., Wagner, A. and Gaudrain, E. (2016). Cognitive compensation of speech perception with hearing impairment, cochlear implants, and aging: how and to what degree can it be achieved?, *Trends Hear.* **20**, 1–16. DOI:10.1177/2331216516670279.

Beauchamp, M. S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex, *Curr. Op. Neurobiol.* **15**, 145–153.

Bergman, M., Blumenfeld, V. G., Cascardo, D., Dash, B., Levitt, H. and Margulies, M. K. (1976). Age-related decrement in hearing for speech: sampling and longitudinal studies, *J. Gerontol.* **31**, 533–538.

Bermant, R. I. and Welch, R. B. (1976). Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition, *Percept. Mot. Skills* **42**, 487–493.

Bertelson, P. and Aschersleben, G. (1998). Automatic visual bias of perceived auditory location, *Psychon. Bull. Rev.* **5**, 482–489.

Bertelson, P., Vroomen, J., de Gelder, B. and Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention, *Percept. Psychophys.* **62**, 321–332.

Bosmana, A. J. and Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment, *Audiology* **34**, 260–284.

Chen, L. and Vroomen, J. (2013). Intersensory binding across space and time: a tutorial review, *Atten. Percept. Psychophys.* **75**, 790–811.

Cienkowski, K. M. and Carney, A. E. (2002). Auditory-visual speech perception and aging, *Ear Hear.* **23**, 439–449.

Colonius, H. and Arndt, P. (2001). A two-stage model for visual–auditory interaction in saccadic latencies, *Percept. Psychophys.* **63**, 126–147.

Couth, S., Gowen, E. and Poliakoff, E. (2018). Using race model violation to explore multi-sensory responses in older adults: enhanced multisensory integration or slower unisensory processing?, *Multisens. Res.* **31**, 151–174.

de Boer-Schellekens, L. and Vroomen, J. (2014). Multisensory integration compensates loss of sensitivity of visual temporal order in the elderly, *Exp. Brain Res.* **232**, 253–262.

de Dieuleveult, A. L., Siemonsma, P. C., van Erp, J. B. F. and Brouwer, A.-M. (2017). Effects of aging in multisensory integration: a systematic review, *Front. Aging Neurosci.* **28**, 80. DOI:10.3389/fnagi.2017.00080.

de Dieuleveult, A. L., Perry, S., Siemonsma, P. C., Brouwer, A. M. and van Erp, J. B. F. (2019). A simple target interception task as test for activities of daily life performance in older adults, *Front. Neurosci.* **13**, 524. DOI:10.3389/fnins.2019.00524.

de Gelder, B. and Bertelson, P. (2003). Multisensory integration, perception and ecological validity, *Tr. Cogn. Sci.* **7**, 460–467.

Diederich, A., Colonius, H. and Schomburg, A. (2008). Assessing age-related multisensory enhancement with the time-window-of-integration model, *Neuropsychologia* **46**, 2556–2562.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading, *Nature* **381**, 66–68.

Erber, N. P. (1975). Auditory-visual perception of speech, *J. Speech Hear. Disrd.* **40**, 481–492.

Ernst, M. O. and Bülthoff, H. H. (2004). Merging the senses into a robust percept, *Trends Cogn. Sci.* **8**, 162–169.

Files, B. T., Tjan, B. S., Jiang, J. and Bernstein, L. E. (2015). Visual speech discrimination and identification of natural and synthetic consonant stimuli, *Front. Psychol.* **6**, 878. DOI:10.3389/fpsyg.2015.00878.

Fogerty, D., Humes, L. E. and Busey, T. A. (2016). Age-related declines in early sensory memory: identification of rapid auditory and visual stimulus sequences, *Front. Aging Neurosci.* **8**, 90. DOI:10.3389/fnagi.2016.00090.

Fozard, J. L. (1990). Vision and hearing in aging, in: *Handbook of the Psychology of Aging*, J. E. Birren and K. W. Schaie (Eds), pp. 150–170. Academic Press, New York, NY, USA.

Freiherr, J., Lundström, J. N., Habel, U. and Reetz, K. (2013). Multisensory integration mechanisms during aging, *Front. Hum. Neurosci.* **7**, 863. DOI:10.3389/fnhum.2013.00863.

Gardner, W. G. and Martin, K. D. (1995). HRTF measurements of a KEMAR, *J. Acoust. Soc. Am.* **97**, 3907. DOI:10.1121/1.412407.

Grant, K. W. and Seitz, P. F. (1998). Measures of auditory–visual integration in nonsense syllables and sentences, *J. Acoust. Soc. Am.* **104**, 2438–2450.

Hay-McCutcheon, M. J., Pisoni, D. B. and Hunt, K. K. (2009). Audiovisual asynchrony detection and speech perception in hearing-impaired listeners with cochlear implants: a preliminary analysis, *Int. J. Audiol.* **48**, 321–333.

Helfer, K. S. (1998). Auditory and auditory–visual recognition of clear and conversational speech by older adults, *J. Am. Acad. Audiol.* **9**, 234–242.

Hoffman, H. J., Dobie, R. A., Ko, C.-W., Themann, C. L. and Murphy, W. J. (2012). Hearing threshold levels at age 70 years (65–74 years) in the unscreened older adult population of the United States, 1959–1962 and 1999–2006, *Ear Hear.* **33**, 437–440.

Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: some statistical considerations, *Brain Topogr.* **21**, 168–176.

Katz, J. (2014). *Handbook of Clinical Audiology*, 5th edn. Lippincott, Williams, and Wilkins, Philadelphia, PA, USA.

Lalonde, K. and Holt, R. F. (2016). Audiovisual speech perception development at varying levels of perceptual processing, *J. Acoust. Soc. Am.* **139**, 1713–1723.

Langendijk, E. H. A. and Bronkhorst, A. W. (2002). Contribution of spectral cues to human sound localization, *J. Acoust. Soc. Am.* **112**, 1583–1596.

Laurienti, P. J., Burdette, J. H., Maldjian, J. A. and Wallace, M. T. (2006). Enhanced multisensory integration in older adults, *Neurobiol. Aging* **27**, 1155–1163.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics, *J. Acoust. Soc. Am.* **49**, 467–477.

Lovelace, C. T., Stein, B. E. and Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection, *Cogn. Brain Res.* **17**, 447–453.

Macpherson, E. A. and Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited, *J. Acoust. Soc. Am.* **111**, 2219–2236.

Mahoney, J. R., Holtzer, R. and Verghese, J. (2014). Visual-somatosensory integration and balance: evidence for psychophysical integrative differences in aging, *Multisens. Res.* **27**, 17–42.

Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., Suzuki, Y., Watanabe, K., Wierstorf, H., Ziegelwanger, H. and Noisternig, M. (2013). Spatially oriented format for acoustics: a data exchange format representing head-related transfer functions, in: *Proceedings of the 134th Convention of the Audio Engineering Society (AES) Roma, Italy*, Convention Paper 8880.

Massaro, D. W., Cohen, M. M. and Smeele, P. M. T. (1996). Perception of asynchronous and conflicting visual and auditory speech, *J. Acoust. Soc. Am.* **100**, 1777–1786.

Mattys, S. L., Davis, M. H., Bradlow, A. R. and Scott, S. K. (2012). Speech recognition in adverse conditions: a review, *Lang. Cogn. Process.* **27**, 953–978.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices, *Nature* **264**, 746–748.

Middlebrooks, J. C. and Green, D. M. (1991). Sound localization by human listeners, *Annu. Rev. Psychol.* **42**, 135–159.

Musacchia, G., Arum, L., Nicol, T., Garstecki, D. and Kraus, N. (2009). Audiovisual deficits in older adults with hearing loss: biological evidence, *Ear Hear.* **30**, 505–514.

Oppenheim, A. V., Schafer, R. W. and Buck, J. R. (1999). *Discrete-Time Signal Processing*, 2nd edn. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Otte, R. J., Agterberg, M. J. H., Van Wanrooij, M. M., Snik, A. F. M. and van Opstal, J. (2013). Age-related hearing loss and ear morphology affect vertical but not horizontal sound-localization performance, *J. Assoc. Res. Otolaryngol.* **14**, 261–273.

Peiffer, A. M., Mozolic, J. L., Hugenschmidt, C. E. and Laurienti, P. J. (2007). Age-related multisensory enhancement in a simple audiovisual detection task, *Neuroreport* **18**, 1077–1081.

Perrott, D. R. and Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth, *J. Acoust. Soc. Am.* **87**, 1728–1731.

Pichora-Fuller, M. K., Schneider, B. A. and Daneman, M. (1995). How young and old adults listen to and remember speech in noise, *J. Acoust. Soc. Am.* **97**, 593–608.

Pick, H. L., Warren, D. H. and Hay, J. C. (1969). Sensory conflict in judgments of spatial direction, *Percept. Psychophys.* **6**, 203–205.

Rosenblum, L. D. and Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception, *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 318–331.

Saija, J. D., Akyürek, E. G., Andringa, T. C. and Başkent, D. (2014). Perceptual restoration of degraded speech is preserved with advancing age, *J. Assoc. Res. Otolaryngol.* **15**, 139–148.

Saija, J. D., Başkent, D., Andringa, T. C. and Akyürek, E. G. (2019). Visual and auditory temporal integration in healthy younger and older adults, *Psychol. Res.* **83**, 951–967.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition, *Psychol. Rev.* **103**, 403–428.

Setti, A., Burke, K. E., Kenny, R. A. and Newell, F. N. (2011). Is inefficient multisensory processing associated with falls in older people?, *Exp. Brain Res.* **209**, 375–384.

Shaw, E. A. G. (1974). Transformation of sound pressure level form the free field to the eardrum in the horizontal plane, *J. Acoust. Soc. Am.* **56**, 1848–1861.

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 Hear.* **26**, 263–275.

Søndergaard, P. L. and Majdak, P. (2013). The auditory modeling toolbox, in: *The Technology of Binaural Listening, Modern Acoustics and Signal Processing*, J. Blauert (Ed.), pp. 397–425. Springer-Verlag, Heidelberg, Berlin, Germany. Available at: http://link.springer.com/10.1007/978-3-642-37762-4.

Stein, B. E. and Meredith, M. A. (1993). *The Merging of the Senses*. MIT Press, Cambridge, MA, USA.

Stevenson, R. A., Nelms, C. E., Baum, S. H., Zurkovsky, L., Barense, M. D., Newhouse, P. A. and Wallace, M. T. (2015). Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition, *Neurobiol. Aging* **36**, 283–291.

Stevenson, R. A. and Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies, *Exp. Brain Res.* **227**, 249–261.

Strouse, A., Ashmead, D. H., Ohde, R. N. and Grantham, D. W. (1998). Temporal processing in the aging auditory system, *J. Acoust. Soc. Am.* **104**, 2385–2399.

Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F. and Théoret, H. (2007). Speech and non-speech audio-visual illusions: a developmental study, *PLoS ONE* **2**(8), e742. DOI:10.1371/journal.pone.0000742.

Tuomainen, J., Andersen, T. S., Tiippana, K. and Sams, M. (2005). Audio-visual speech perception is special, *Cognition* **96**, B13–B22.

Tye-Murray, N., Sommers, M. S. and Spehar, B. (2007). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing, *Ear Hear.* **28**, 656–668.

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, *Int. J. Audiol.* **47**, S31–S37.

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J. and Hale, S. (2010). Aging, audiovisual integration, and the principle of inverse effectiveness, *Ear Hear.* **31**, 636–644.

Van Opstal, A. J. (2016). *The Auditory System and Human Sound-Localization Behavior*. Academic Press, London, UK.

Vatakis, A. and Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions, *Brain Res.* **1111**, 134–142.

Vroomen, J. and de Gelder, B. (2004). Perceptual effects of cross-modal stimulation: ventriloquism and the freezing phenomenon, in: *The Handbook of Multisensory Processes*, G. A. Calvert, C. Spence and B. E. Stein (Eds), pp. 141–150. The MIT Press, Cambridge, MA, USA.

Vroomen, J., Bertelson, P. and de Gelder, B. (1998). A visual influence in the discrimination of auditory location, in: *Proceedings of the International Conference on Auditory–Visual Speech Processing (AVSP'98)*, D. Burnham, J. Robert-Ribes and E. Vatikiotis-Bateson (Eds), pp. 131–134. Causal Productions, Adelaide, South Australia, Australia.

Vroomen, J., Bertelson, P. and de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention, *Percept. Psychophys.* **63**, 651–659.

Wenzel, E. M., Arruda, M., Kistler, D. J. and Wightman, F. L. (1993). Localization using non-individualized head-related transfer functions, *J. Acoust. Soc. Am.* **94**, 111–123.

Wightman, F. L. and Kistler, D. J. (1989). Headphone simulation of free-field listening. II: psychophysical validation, *J. Acoust. Soc. Am.* **85**, 868–878.

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H. and Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention, *J. Neurosci.* **32**, 14010–14021.

Ziegelwanger, H. and Majdak, P. (2014). Modeling the direction-continuous time-of-arrival in head-related transfer functions, *J. Acoust. Soc. Am.* **135**, 1278–1293.