UNIVERZITA PAVLA JOZEFA ŠAFÁRIKA V KOŠICIACH
PRÍRODOVEDECKÁ FAKULTA

MODELING OF THE REFERENCE FRAME OF THE VENTRILOQUISM AFTEREFFECT

2021 Ing. Peter LOKŠA

UNIVERZITA PAVLA JOZEFA ŠAFÁRIKA V KOŠICIACH

PRÍRODOVEDECKÁ FAKULTA

MODELING OF THE REFERENCE FRAME OF THE VENTRILOQUISM AFTEREFFECT

DIZERTAČNÁ PRÁCA

Študijný program:	Informatika
Pracovisko (katedra/ústav):	Ústav informatiky
Vedúci diplomovej práce:	doc. Ing. Norbert Kopčo, PhD.

Košice 2021

Ing. Peter LOKŠA

Univerzita P. J. Šafárika v Košiciach
Prírodovedecká fakulta

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Ing. Peter Lokša
Študijný program: Informatika (Jednoodborové štúdium, doktorandské III. st., denná forma)
Študijný odbor: Informatika
Typ záverečnej práce: Dizertačná práca
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Modeling of the Reference Frame of the Ventriloquism Aftereffect

Názov SK: Modelovanie vzťažnej sústavy bruchomluveckého afterefektu

Cieľ: Experiments and modeling examining how spatial information is integrated across the auditory and visual modality in spatial processing.

Literatúra: Kopčo, N, I-F Lin, BG Shinn-Cunningham, and JM Groh (2009). Reference frame of the ventriloquism aftereffect. Journal of Neuroscience, 29(44):13809-13814
Kopčo N, Lin, I-F. Shinn-Cunningham B, Groh JM (2008). "Visual calibration of auditory spatial perception in humans and monkeys," Presented at the 31st MidWinter meeting of the Association for Research in Otolaryngology, Phoenix, Arizona, USA, February 16-21, 2008.

Kľúčové slová: computational modeling, reference frame, ventriloquism aftereffect, auditory saccades, auditory spatial representation

Školiteľ: doc. Ing. Norbert Kopčo, PhD.
Ústav : ÚINF - Ústav informatiky
Riaditeľ ústavu: doc. RNDr. Ondrej Krídlo, PhD.

Dátum schválenia: 01.09.2013

Abstrakt v štátnom jazyku

Mozog používa rôzne metódy na zakódovanie pozícií zrakových verzus sluchových podnetov. Sietnica sníma pozície zrakových objektov s ohľadom na oči, ktoré sa hýbu, keď sa obzeráme. Sluchový systém získava priestorovú informáciu z rozdielov v charakteristikách zvuku pre jedno verzus druhé ucho, čo ukazuje, že pozície zdrojov zvuku sú vztiahnuté na hlavu. Je potrebné, aby tieto dve vzťažné sústavy boli zarovnané, aby sa vytvoril koherentny sluchovo-zrakový priestorový vnem, alebo aby sa mohla uskutočniť zrakom riadená rekalibrácia sluchového priestoru. Cieľom tejto práce je preskúmať a modelovať procesy zarovnania súradnicových sústav reprezentácie zraku a sluchu v mozgu. Známym javom, v ktorom zrak a sluch interagujú, je „bruchomluvecký afterefekt“ (VA z angl. ventriloquism aftereffect), ktorý je pozorovávaný ako posun vo vnímaných pozíciách zvukuv po opakovaných sluchovo-zrakových podnetoch s nezarovnanými sluchovými a zrakovými zložkami. Predchádzajúca štúdia ukázala že vzťažná sústava bruchomluveckého afterefektu je zmesou vzťažných sústav orientovaných na polohu hlavy (HC – angl. head-centered) a polohu očí (EC – angl. eye-centered), kde VA je vyvolaný v strede sluchovo-zrakového poľa. Po prvé, v tejto dizertačnej práci boli analyzované výsledky nového experimentu, v ktorom bol tento efekt vytvorený v sluchovo-zrakovej periférii, aby sa preskúmalo, či uniformity v kódovaní audiovizuálneho priestoru ovplyvňujú pozorovanú vzťažnú sústavu. V periférií bola vzťažná sústava identifikovaná ako primárne centrovaná na polohu hlavy. Výsledky tiež ukázali novú formu zrakom spôsobenej adaptácie v sluchovej reprezentácií, ktorá závisela na informáciách centrovaných na polohu aj hlavy aj očí. Po druhé, bol vyvinutý výpočtový model aby sa preskúmali pozorované adaptácie a transformácie, so zameraním na dáta pre vzťažnú sústavu z popísaných dvoch experimentálnych štúdií. Boli vykonané dve predbežné a jedna hlavná modelovacia štúdia. V každej z nich boli navrhnuté, fitované a testované rôzne verzie modelu. Finálny model mal dve hlavné verzie, z ktorých každá obsahovala dve aditívne skombinovné zložky: sakadická (týkajúca sa pohybov očí) zložka charakterizujúca adaptáciu sluchových sakadických odpovedí, a zložka pre priestorovo-sluchovú reprezentáciu adaptovanú bruchomluveckými signálmi. Obe verzie modelu boli evaluované a odlišovali sa v tom, či boli signály vo vzťažnej sústave centrované na polohu hlavy (HC verzia), alebo boli kombináciou vzťažných sústav centrovaných na polohy hlavy a očí (HEC verzia). HEC model mal lepšie výsledky v porovnaní s modelom HC v hlavnej simulácií, ktorá brala do úvahy všetky dáta, kým HC model bol vhodný keď bola adaptácia vyvolávaná len zarovnanými sluchovo-zrakovými stimulmi. Tieto výsledky podporujú závery, že kým je VA ovplyvňovaný viacerými priestorovými nerovnomernými hemisfericky-špecfickými procesmi, na kalibráciu sluchovej priestorovej reprezentácie sú použité vizuálne signály v uniformnej zmiešanej HC+EC vzťažnej sústave, dokonca aj po zohľadnení EC sluchovej sakadickej adaptácie.

Kľúčové slová: výpočtové modelovanie, vzťažná sústava, bruchomluvecký afterefekt, sluchové sakády, priestorovo-sluchová reprezentácia

Abstrakt v cudzom jazyku

The brain uses different methods for encoding the locations of visual versus auditory stimuli. The retina senses the locations of visual objects with respect to the eyes, which move as we look around. The auditory system extracts spatial information from the differences in sound characteristics across the ears, indicating the locations of sound sources referenced to the head. These two reference frames need to be aligned to create a coherent audiovisual spatial percept, or to induce a visually guided recalibration of the auditory space. The goal of the current thesis is to examine and model the process of alignment of the coordinate frames of representation of vision and hearing in the brain. A well-known phenomenon in which vision and hearing interact is the ventriloquism aftereffect (VA), observed as a shift in the perceived locations of sounds after repeated audiovisual stimulation with misaligned auditory and visual components. A previous study showed that the reference frame (RF) of the VA is a mixture of head-centered (HC) and eye-centered (EC) RFs when the VA is induced in the center of audiovisual field. Here, first, the results of an experiment were analyzed in which the effect was induced in the audiovisual periphery to examine whether uniformities in the encoding of auditory space affect the observed reference frame. Indeed, in the periphery, the reference frame is found to be predominantly head-centered. Also, the results showed a new form of visually induced adaptation in the auditory representation, dependent on both eye-centered and head-centered spatial information. Second, a computational model is developed to examine the observed adaptations and transformations, focusing on the reference frame data from the two experimental studies. Two preliminary and one main modeling study were performed. In each of them, different versions of the model were designed, fitted, and tested. The final model has two main versions, each containing two additively combined components: a saccade-related component characterizing the adaptation in auditory-saccade responses, and auditory space representation adapted by the ventriloquism signals. Two versions of the model are evaluated, differing by whether the ventriloquism signals are in the HC RF (HC version) or in a combination of HC and EC RFs (HEC version). The HEC model performed better than the HC model in the main simulation considering all the data, while the HC model was more appropriate when only the AV-aligned adaptation data were simulated. These results support the conclusions that, while the ventriloquism aftereffect is driven by multiple spatially non-uniform, hemisphere-specific processes, visual signals in a uniform mixed HC+EC RF are likely used to calibrate the auditory spatial representation, even after the EC-referenced auditory-saccade adaptation is accounted for.

Keywords: computational modeling, reference frame, ventriloquism aftereffect, auditory saccades, auditory spatial representation

Table of contents

List of abbreviations

VE	ventriloquism effect
VA	ventriloquism aftereffect
AV	audiovisual
RF	reference frame

Introduction

Spatial hearing is plastic. It can adapt itself in response to many factors. E.g., the adaptation can be induced by visual information, as manifested by ventriloqusim effect or ventriloqusim aftereffect.

The ventrilqusim effect is an immediate mis-callibration in sound localization induced by a synchronously presented visual stimulus located at a location slightly displaced from the sound location. The ventriloquism effect is exhibited by a shift in the perceived auditory stimulus location toward the visual stimulus which can be as large as ~80% of the audio-visual discrepancy.

The ventriloquism aftereffect is an effect which endures for a longer time (from seconds to minutes) after the presentation of the misaligned audio-visual stimuli. Usually it is induced by multiple repeated presentations of the audio-visual stimuli, and the result is a mis-localization of a sound presented alone, which is perceptually shifted towards the visual component of the previously presented audio-visual stimuli. The strength of the aftereffect is typically smaller than the strength of the effect.

The the reference frame of spatial vision is known to be eye-centered, while the reference frame of spatial hearing is head-centered. In order to integrate these two percepts into a single audiovisiual percept, the brain needs to align these reference frames into a common reference frame.

A previous experiment (Kopco *et al.*, 2009) was performed to identify the audiovisual reference frame of the ventriloquism aftereffect in humans and non-human primates. It examined the ventriloquism aftereffect to see its generalization to both candidate reference frames when the eyes shifted to a new location. Its results suggested that the reference frame of ventrilqusim aftereffect could be a mixture of head- and eye-centered reference frames. The limitation of that study was that the ventriloquism aftereffect was inusted only in the center of the visual field. Thus, it was not clear whether the same pattern of results would be observed in other parts of the audiovisual field.

In the first part of this thesis (chapter 3), we describe the results of analysis of a new study designed to examine the reference frame if the aftereffect is induced in the visual periphery. The results were inconsistent with the previous findings, suggesting the head-centered reference frame of ventriloquism aftereffect to be prevalent in this part of the visual field. Then, we designed, implemented, and tested a model describing these experimental data, described in the second part of this thesis (Chapters 4, 5 and 6).

1 State of the art

1.1 Cross-modal interactions and audio-visual cross-modal illusions

The individual senses interact to a far greater extent than one might expect. And there is a broad range of cross-modal interactions, many of which are without any obvious purpose. E.g., early studies observed that tactile object recognition is weakened in darkness (Johnson, 1920), or that perception of touch or pressure is improved by the presence of quiet sounds while being degraded by the presence of the loud sounds (Urbantschitsch, 1888) .

There are several main types of cross-modal interactions: enhancement, effect of intensity, habituation, and masking (Stein and Meredith, 1993; Calvert *et al.*, 2004; Kopco, 2020). The cross-modal enhancement is observed as facilitation of the primary percept by the weak secondary percept. However, the relative intensity of the two percepts is also a factor, such that the secondary percept can cause inhibition of the primary percept if the secondary percept is very strong. The habituation resides in decreasing of the sensitivity to the secondary stimulus due to its repetition without perceivable meaningful event after it. Masking is observed as inhibition of the primary stimulus when the secondary stimulus is located at a distinct but nearby location. The enhancement and masking are therefore similar, with the key difference determining the positive effect (enhancement) from the negative effect (masking) residing in the relative location of the two stimuli.

The strength with which one modality influences another modality depends on many factors, dominant among which is the accuracy with which the different modalities encode the information important for a given task, or task domain (Klatzky and Lederman, 2010). The modality appropriateness hypothesis suggests that the cross-modal influence of any given modality in a given task domain grows with the modality accuracy in that domain. In spatial domain, vision dominates over hearing, as vision is orders of magnitude more accurate in spatial encoding than hearing. Conversely, in the time domain, hearing dominates over vision, which can be predicted by the observation that hearing encodes temporal information much more accurately than vision.

Several cross-modal illusions have been particularly impactful in the cognitive neuroscience research. The McGurk effect (McGurk and MacDonald, 1976) is an illustration of audio-visual interactions in speech perception. In this illusion, when watching a video of a person speaking a specific syllable (e.g., “ga”) while listening to a synchronized audio recording of the same person saying a different syllable (e.g., “ba”), the resulting percept is often a syllable different from either the visual or the auditory stimulus (e.g., “da” or “tha”). This illusion is important because it supports the “motor theory” of speech perception, which suggests that, when listening to speech, the listener’s brain “simulates” the function of speech organs of the speaker in order to identify the content of the speech (Liberman and Mattingly, 1985).

Double flash illusion (Roseboom *et al.*, 2013) is an illusion that illustrates the dominance of the auditory modality in the temporal domain. In the illusion, the subject is presented with one visual stimulus accompanied by two auditory beeps. A naive subject typically has the impression that there were two flashes, thus illustrating that the temporal information from the auditory modality interferes with and dominates over the information from the visual domain.

Finally, the Ventriloquism effect (Howard and Templeton, 1966; Jack and Thurlow, 1973; Choe *et al.*, 1975; Vroomen *et al.*, 2001; Alais and Burr, 2004; Hendrickx *et al.*, 2015; Park and Kayeser, 2020; Tong *et al.*, 2020) is a spatial perceptual audiovisual illusion. It is exhibited as a shift of the localization of the sound by a spatially displaced synchronous visual stimulus. The ventriloquism effect, which is immediate, can result in ventriloquism aftereffect (Canon, 1970; Woods and Recanzone, 2004; Bertelson *et al.*, 2006; Kopco *et al.*, 2009; Bruns *et al.*, 2020), which is induced by repeated presentation of such audio-visual stimuli and which is exhibited by persisting shift in the perceived localization of auditory stimuli even when they are presented alone. The ventriloquism aftereffect illusion is the primary illusion examined here, and the existing models of the ventriloquism effect and aftereffect, which will be summarized in chapter 1.4.

1.2 Spatial auditory processing

Spatial hearing serves us by providing us information about the location of important events in the space (Ahveninen *et al.*, 2014). It uses the information from both of our ears (binaural – related to both ears) to extract the basic spatial information.

Horizontal sound localization is based on information provided by two binaural cues: interaural level difference (ILD) and interaural time difference (ITD) (Feddersen *et al.*, 1957; Middlebrooks and Green, 1991; Ahveninen *et al.*, 2014).

Considering the ILD, for each direction of approaching sound, there is particular difference in sound level for one ear vs. the other. Based on this, specific parts of our brain are able to identify this direction. Analogically, considering the ITD, for each direction of the approaching sound, there is a particular difference in the time of sound arriving at our ears, so our brain can use this auditory localization cue too.

These two cues are, roughly speaking, not complementary in the domain of space, because they give the same spatial information, but they are complementary in the frequency domain, because for the frequencies for which one cue does not work, the other one does. The ILD dominates at the high frequencies (larger than 3 kHz) as the head creates an acoustic shadow which is stronger at higher frequencies. The ITD dominates at the low frequencies (smaller than 1.5 kHz) as the fine structure of the sounds at high frequencies is not detectable by the auditory neurons due to their limited temporal resolution. This complementarity is referred to as the Duplex theory (Rayleigh, 1907; Feddersen *et al.*, 1957).

But because the ITD and ILD are one-dimensional cues and the physical space is three-dimensional, the information provided by these cues is ambiguous, creating the cone of confusion, or a torus of confusion (Shinn-Cunningham *et al.*, 2000). So for a given ITD or ILD value, there are many locations in space which can be matched to this value, and the resulting set made of possible locations is shaped to the surface of a cone (or a torus). One effective method to break such confusion is to move the head in order to collapse the possible space of locations into one concrete point.

While the ILD and ITD are helpful when identifying the direction of the coming sound, they are relatively unable in identifying the distance of the coming sound. One strategy which our auditory system uses to perform such identification of the distance of the coming sound is based on the comparison of the magnitude of the direct sound with the magnitude of the sound reflected from the objects in the environment (i.e. reverberant sound) (Larsen *et al.*, 2008; Kopco and Shinn-Cunningham, 2011; Ahveninen *et al.*, 2014). The reverberant level is roughly fixed for a fixed sound. Thus, the closer the sound source is, the greater the level of the direct sound is present in the sensed sound in comparison with the reverberant level.

The third dimension that needs to be represented is the sound source elevation. The main sounce of information for localization in this dimension is the spectral pinna cue. Its name is derived from the body part of which shape is essential in this identification, because it modifies the incoming sound’s spectrum according to its elevation. Specifically, such identification happens according to the spectral notch due to correlation of its frequency with elevation of the coming sound (Musicant and Butler, 1985; Wightman and Kistler, 1997; Ahveninen *et al.*, 2014).

1.3 Ventriloquism effect, aftereffect, and their reference frames

1.3.1 Ventriloquism effect

The ventriloquism effect is a mislocalization of sound due to a spatially displaced but close and synchronous visual stimulus (Howard and Templeton, 1966; Jack and Thurlow, 1973; Choe *et al.*, 1975; Vroomen *et al.*, 2001; Alais and Burr, 2004; Hendrickx *et al.*, 2015; Park and Kayeser, 2020; Tong *et al.*, 2020). The ventriloquism effect is immediate and its magnitude is around 80% of the localization error relative to the offset of the visual stimulus component from the auditory component (e.g., a 5° visual shift causes a 4° shift in the perceived location of the auditory component of the audiovisual stimulus) (Kopco *et al.*, 2009).

1.3.2 Ventriloquism aftereffect

If the pair of spatially discrepant stimuli for hearing and vision are repeated with constant across-repetition locations for both stimuli, so the discrepancy is constant too, the auditory spatial representation becomes recalibrated which is manifested by the localization error for sounds presented alone in the vicinity of the previously presented audiovisual stimuli. This is called ventriloquism aftereffect (Canon, 1970; Woods and Recanzone, 2004; Bertelson *et al.*, 2006; Kopco *et al.*, 2009; Bruns *et al.*, 2020). The aftereffect is typically weaker than the effect. E.g., localization error was around 25% of the discrepancy when the aftereffect was induced at several locations spanning 20° in the frontal audiovisual field (Kopco *et al.*, 2009). The time course of the ventriloquism aftereffect to build up is on the order of seconds, minutes to tens of minutes (Kopco *et al.*, 2009).

1.3.3 Reference frame of the ventriloquism aftereffect

The reference frame of the spatial hearing is head-centered while the reference frame of the spatial vision is eye-centered (Brainard and Knudsen, 1995; Razavi *et al.*, 2007). The property of spatial hearing being head-centered is given by the fact that the ears are immobile relative to the head, so the stimuli are received at the two ears with physical characteristics that are constant with respect to the head orientation. The property of spatial vision being eye-centered is given by the fact that photosensitive cells are immobile relative to the eye, so the most straightforward way to process the visual stimuli is a direct mapping which is eye-centered.

The brain needs to transform at least one of the unimodal (auditory or visual) spatial information representations to the reference frame of the other modality in order to correctly induce spatial auditory adaptation by visual signals. There were several studies conducted that were researching this topic (Bulkin and Groh, 2012; Lee and Groh, 2012; Van Barneveld and Van Wandrooij, 2013; Mohl *et al.*, 2020).

To examine the reference frame of the ventriloquism aftereffect, Kopco et al. (2009) performed an experiment in which ventriloquism was induced in 7 human subjects and 2 monkeys. The key element of the method was that the authors induced the aftereffect locally while the eyes fixated a single fixation location, and then examined how the pattern of adaptation changed if the eyes are shifted to a new fixation location. It was hypothesized that if the region of the observed adaptation shifted with the eye fixation, it would be an evidence that the reference frame of the ventriloquism aftereffect is eye-centered. On the other hand, if the pattern of adaptation did not change as the eyes moved, that would be considered to be an evidence that the reference frame is head-centered.

The resulting reference frame for both humans and monkeys was a mixture of the two hypothesized reference frames, which was apparent from the results shown in that study.

However, one limitation of the study resided in the spatial area in which the ventriloquism aftereffect was induced, which was in the center of the visual field. In the current thesis, the results of an experiment are first presented in which it was examined what happens if the aftereffect is induced in peripheral part of the visual field (Chapter 3; published in Kopco *et al.* (2019b)). The main result of the experiment is that a predominantly head-centered reference frame is observed, inconsistent with the former finding. The main goal of the modeling presented in Chapters 4-6 is then to examine this inconsistency between the two studies and to provide a uniform model of the reference frame of the ventriloquism aftereffect.

1.3.4 General characteristics and factors influencing the ventriloquism aftereffect

Recent studies described various specific properties of the ventriloquism aftereffect. Here, a summary of several relevant studies is provided.

There are two hypothetical mechanisms through which the visual adaptation/calibration of auditory spatial perception (i.e., the ventriloquism aftereffect) might occur (Pages and Groh, 2013). The calibration might be driven by seeing the object or event while hearing the same object, in which case a synchrony between the auditory and visual stimuli is important. Or, the calibration might be driven by seeing the object or event after hearing it, in which case the visual stimulus provides feedback after the auditory stimulus was presented. Pages & Groh (Pages and Groh, 2013) observed the strongest adaptation in the feedback condition. And, the feedback alone had even greater effect than synchrony and feedback together, providing a strong evidence that the ventriloquism adaptation is mainly driven by feedback visual signals.

Ventriloquism aftereffect is usually built up over the time course of minutes to tens of minutes (Razavi *et al.*, 2007; Brown *et al.*, 2012; Bosen *et al.*, 2018) by repetition of auditory stimulus with spatially discrepant visual adaptor. However, under specific conditions, such adaptation can be triggered by single audio-visual trials over the time course ranging from milliseconds to seconds (Wozny and Shams, 2011). Also, the Wozny and Shams study (Wozny and Shams, 2011) observed a direct relationship between the size of mislocalization on an auditory tiral and the discrepancy of visual re. auditory stimulus in the previous audiovisual trial, again supporting the notion that there is a very fast component to the ventriloquism adaptation.

Maddox *et al.* (2014) showed that gaze direction can affect not only the perceived sound location, but also the listener's ability to discriminate the location of sounds. The subjects in that study were exposed to pairs of nearby auditory stimuli (generated by either manipulation the stimulus ITD or its ILD), preceded by visual or auditory primer which could be informative or uninformative. Moving the eyes to fixate the visual primer had a significant effect on the percentage of correct discrimination in case of both ITD and ILD.

These results are consistent with shifting spatial receptive field for which the response rate is more different for auditory stimuli in the center of the eye-centered visual field than in its periphery for identical displacement of these two auditory stimuli, assuming that the response rate as the function of stimulus azimuth is sigmoid.

Typically, the ventriloquism aftereffect is induced by visual stimuli that are offset by a constant angular amount from the auditory stimuli, resulting in a constant shift in the auditory localization percepts. Zwiers et al. (2003) showed that if the subject wears glases with lenses compressing the visual field, the resulting auditory spatial representation is also compressed. One of the research questions in this experiment was what happens with the parts of audiovisual spatial map outside the range visible through the lenses after wearing these lenses for a sufficient amount of time (2 or 3 days in this case). The resulting behavior for the sound coming from these blocked parts of the visual field are that the perceived re. physical locations were shifted by a constant amount, meaning that all locations shifted by the same angle, corresponding to the amount by which the lenses shifted the vision at the outer edge of the visible area.

1.4 Models of the ventriloquism aftereffect and effect

This chapter provides an overview of the existing models of audiovisual crossmodal integration and visually induced spatial auditory adaptation.

1.4.1 Models of the ventriloquism effect

Alais and Burr (2004), Mendonca *et al.* (2016) and Odegaard *et al.* (2016) presented causal inference models describing how the auditory stimulus perception is biased by the synchronous visual stimulus. According to these models our brain first decides whether to categorize the auditory and the visual stimulus as corresponding to a single event vs. two separate events. In the former case, a bimodal spatial map is used in localization, so a new estimated location is computed based on both auditory and visual location estimates. In the latter case, the auditory stimulus is not significantly biased by visual information, as only the auditory location estimate is used.

According to these models, to decide whether there is a single event or two events, the brain uses probability estimation using a Gaussian function of the physical stimulus source location for each modality and it computes the analogical bimodal version of this function. The peak of the unimodal auditory function is then compared to the peak of the bimodal one, and consequently, it chooses the greater value, and finally, the resulting azimuth is evaluated and used as the perceived location of the auditory stimulus.

Importantly, the causal inference models assume that the probability of the two events option is very large. Contrary to that, the data and models in the current thesis were obtained with a relatively small audio-visual displacement of 5° for which all the audiovisual stimuli are generally expected to be integrated.

1.4.2 Models of the ventriloquism aftereffect

This section provides an overview of the models of the ventriloquism aftereffect. These models differ from each other mainly by the specific characteristic of the aftereffect that they consider.

Watson *et al.* (2019) tested three models of ventriloquism aftereffect which have taken the time variable into account. For all of these models the ventriloquism aftereffect shift converged to constant positive percentage less than 100% when the audiovisual stimulus was being presented, and converged to zero percent when no audiovisual stimulus was being presented. The differences between models resided in the mechanisms of the convergence.

For the first tried model of Watson *et al.* (2019) the convergence was implemented as exponential function, for the second model as two exponential functions with time constants and other settings different for each function, and for the third model, it was implemented as the power function. The Akaike information criterion (AIC) was used to compare the models. According to this criterion, the model of two exponential functions with two different time constants was the most appropriate, consistent with the idea that the visually induced adaptation might be occurring in two different neural structures, one relatively fast and one relatively slow.

Bosen *et al.* (2018) proposed a model of the aftereffect consisting of several components, each focused on a different parameter influencing the ventriloquism aftereffect. The main components were the spatial window, which could be flat or triangular, and the decay function, which could be a single exponential, a double exponential, or a power function. Then they tested various model versions, each considering various combinations of the components.

The spatial window function characterizes how the strength of the ventriloquism aftereffect decreases with the azimuthal separation of an auditory-only target from the auditory component of the ventriloquism-inducing audiovisual stimulus. The decay function was describing the temporal dynamics of the ventriloquism aftereffect similarly to Watson *et al.* (2019). The best fitting model according to corrected Akaike information criterion (AICc) was the one with spatial window and power decay function. While this model takes both the spatial and the temporal properties of the ventriloquism aftereffect into account, it still did not consider the different reference frames of the two modalities.

The model of Shinn-Cunningham (2000) is based on optimum decision theory and it was used ot model auditory spatial adaptation data in which shift in responses could be induced by vision or by other forms of feedback. It predicts the response bias as a function of the difference between optimal criterion and the decision-axis criterion in the numerator and the square root of the linear function of the effective stimulus range in the denominator. The optimal criterion is the average of the internal representations of the remapped versions of the current and the previous stimulus. The decision axis criterion is the average of the underlying decision spaces of the current and the previous stimulus. The effective stimulus range is equal to the differences of the maximum and the minimum stimulus azimuth both divided by the current slope. The slope evolves with the time. Neither this model nor the models mentioned earlier considered the reference frame of the visual and auditory representations, and their implications for the adaptation.

Pouget *et al.* (2002) proposed a model of the ventriloquism effect, or of other multimodal phenomena that require an instantaneous alignment of the sensory representations across multiple modalities, that considers the reference frame differences between modalities. The model considers the behavior of neurons as the basic elements of the reference frame alignment. One of the basic ideas of the model is in an ordering of the audiovisual neurons in a 2-dimensional matrix for which each neuron is virtually assigned to particular pair of visual location (the 1st dimension) and oculomotor position (the 2nd dimension). Thus each of these neurons (and therefore also the particular pair) is associable with particular auditory location (according to the rule that the auditory location is equal to the visual location minus oculomotor position). And when the ventriloquism effect is induced, the network of such neurons adapts its weights to correctly predict the relation between the auditory and visual stimuli on the one side and the oculomotor response on the other.

1.4.3 Summary

While the models summarized here considered various aspects of the ventriloquism effect and aftereffect, none of them considered the reference frame of the ventriloquism aftereffect, the main topic of the current dissertation described in Chapters 4-6.

2 Goals and structure

The main goal of the work described in this thesis was to elucidate the nature of the audiovisual spatial calibration, specifically the ventriloquism effect, with respect to its reference frame (Kopco *et al.*, 2009).

In order to achieve this goal, the following steps were performed:

- experimental data of a new behavioral experiment were analyzed (described in chapter 3). These results have been published in Kopco *et al.* (2019b) and in the preprint Kopco *et al.* (2019a).
- a model was developed and evaluated (described in chapter 6). The design of the model is described in section 6.4. The evaluation of the model fitness (i.e. testing) was done in the section 6.6. These results are published in the preprint Loksa and Kopco (2021) and have been submitted for a journal publication.

In addition, several preliminary models were initially considered. Two main ones are described in chapters 4 and 5. These preliminary versions have been published in conference proceedings Loksa and Kopco (2016); 2017) .

Each of the following chapters is a self-contained publication that defines its own goals and describes the work performed to achieve them.

3 Hemisphere-specific properties of the ventriloquism aftereffect

This chapter contains the content of the article Kopco *et al.* (2019b).

The data for the analyses in this article were collected by the authors of Kopco *et al.* (2009). The analysis, interpretation and presentation of the results were performed by the author of this thesis, under supervision of the thesis advisor.

3.1 ABSTRACT

Visual calibration of auditory space requires re-alignment of representations differing in 1) format (auditory hemispheric channels vs. visual maps) and 2) reference frames (head-centered vs. eye-centered). Here, a ventriloquism paradigm from Kopčo *et al.* (J Neurosci, 29, 13809-13814) was used to examine these processes in humans for ventriloquism induced within one spatial hemifield. Results show that 1) the auditory representation can be adapted even by aligned audio-visual stimuli, and 2) the spatial reference frame is primarily head-centered, with a weak eye-centered modulation. These results support the view that the ventriloquism aftereffect is driven by multiple spatially non-uniform, hemisphere-specific processes.

3.2 Introduction

Vision plays an important role in calibration of auditory spatial perception. In the “ventriloquism aftereffect” (VAE), repeated presentations of spatially mismatched visual and auditory stimuli produce a shift in perceived sound location that persists when the sound is presented alone (Canon, 1970; Recanzone, 1998a; Woods and Recanzone, 2004; Bertelson *et al.*, 2006). The brain mechanisms that support this process are mysterious because spatial representations seem to differ in vision and in hearing in two ways.

First, visual space is initially encoded relative to the direction of the eye gaze, while the cues for auditory space are first computed relative to the orientation of the head (Groh & Sparks, 1992). A means of reconciling this discrepancy in reference frames (RF) is necessary to achieve correct recalibration. Our previous study suggests that a mixture of eye-centered and head-centered RFs are associated with recalibration in the central region of the audiovisual field (Kopco *et al.*, 2009).

Second, there is growing evidence that, in mammals, auditory space is encoded non-homogeneously, based on two (or more) spatial channels roughly aligned with the left and right hemifields of the horizontal plane (Grothe *et al.*, 2010; Groh, 2014). This is markedly different from visual spatial codes, in which the retinal surface provides a map of the position of stimuli in the environment.

Thus, the process of using visual information to recalibrate auditory space is multifaceted, and may operate differently in different portions of the environmental scene. Indeed, differential patterns of adaptation across auditory space have been observed (Phillips and Hall, 2005; Maier *et al.*, 2010), suggesting that the auditory code in humans likely employs the same two-channel scheme that has been observed in animal species (Salminen *et al.*, 2009).

Here, we tested whether the spatial characteristics of the ventriloquism aftereffect induced in the audiovisual periphery (i.e., in a single hemifield) differ from those occurring when the aftereffect is induced in the central region (i.e., covering both hemifields; Kopco *et al.*, 2009). Persistent visually driven biases in perceived sound location were induced. As in Kopco *et al.* (2009), we presented mismatched (5°-shifted) audio-visual (AV) stimuli in only a subregion of space (Figure 1A, top panel), but this time the training region was peripheral, rather than central, to the fixation point used for these trials. We evaluated the effects of this pairing on saccade accuracy for interleaved auditory-only trials both from that fixation point and a non-training fixation point in the opposite hemifield (Figure 1A, bottom panel).

As was the case for our previous study involving central training, the pairing of a displaced visual stimulus induced a local aftereffect in the peripheral trained region. Contrary to the previous study, this aftereffect appeared to be mostly in the head-centered reference frame, as the contribution of an eye-centered component was not readily apparent. However, we also observed biases related to the location of the fixation point, even when the AV stimuli were aligned. Together, these findings confirm the contribution of multiple signals related to different reference frames and representational formats across the horizontal space.

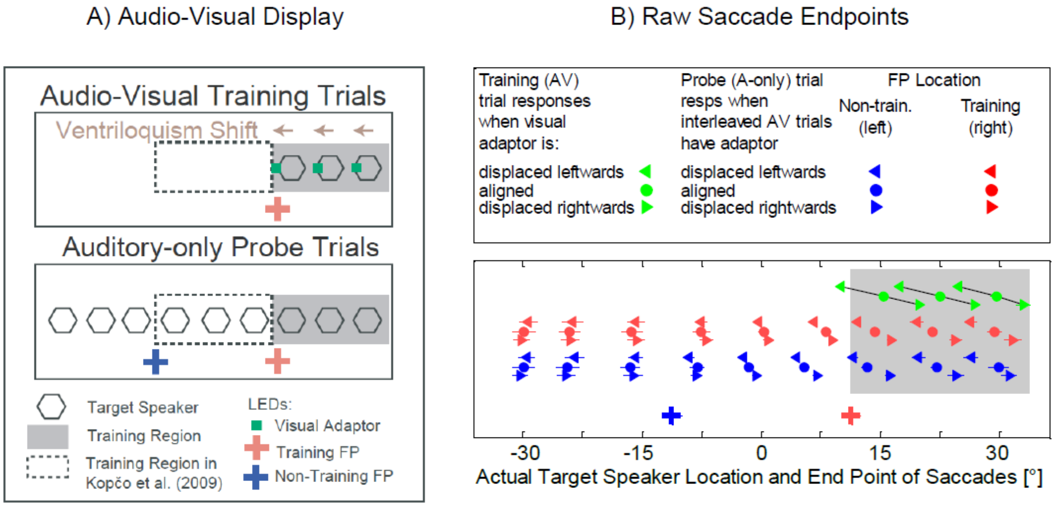

Figure 1: Experimental set-up and raw experimental data. A) Audiovisual display used to present the AV training stimuli in one experimental block. At the beginning of each AV training trial (top), the subject had to fixate on the same initial fixation point (FP); then, the training stimulus was presented from one of three locations lateral to the FP, keeping the direction of the induced shift the same within a block (by consistently presenting the visual adaptor displaced to the left, to the right, or aligned with the target speaker). On the auditory-only probe trials (bottom), the same nine speaker locations and two FPs were used in all blocks. The probe trials were randomly interleaved among the training trials and the FP and target locations varied randomly from trial to trial. Dashed frame indicates the central training region used in Kopco et al. (2009). B) Raw saccade endpoints of the responses to the AV training stimuli and auditory-only probe stimuli as a function of the actual target speaker location, collapsed across time. The symbols represent across-subject mean responses (+/-1 SEM indicated by horizontal lines) in different audiovisual conditions (see legend), separately for the training trials (green), probe trials starting at the training fixation (red), and probe trials starting at the non-training fixation (blue). Graphs for each measurement type are plotted in one row, vertically offset from data for other types, for visual clarity. The A-only data corresponding to each target location are approximately aligned with that target location. For the AV data, the dashed lines connect symbol triplets for the same auditory target when presented with one of the three different visual adaptors (the AV-aligned data are located approximately at the corresponding target location).

3.3 Methods

All procedures and equipment closely matched those used in Kopco *et al.* (2009).

General methods. Experiments were performed in an experimental lab in the Boston University Hearing Research Center. Subjects made eye movements from a visual fixation point to a broadband noise delivered from loudspeakers in darkness. On training trials (Figure 1A, top), visual stimuli were presented simultaneously with the sounds, using light-emitting diodes (LEDs) displaced from the locations of the speakers or aligned with them. On randomly interleaved probe trials (Figure 1A, bottom), only the auditory stimuli were presented.

Subjects. Seven young adults with normal hearing by self-report participated. The experimental protocols were approved by the Boston University institutional review committee.

Setup. Subjects were seated in a quiet darkened experimental room in front of an array of speakers and LEDs (Figure 1). To keep the head-centered RF fixed, the subjects' heads were restrained by a chin rest. Subjects' behavior was monitored and responses were collected by an infrared eye tracker, calibrated using visually guided saccades to selected target locations at the beginning of each session.

Stimuli. Sounds were 100-ms broadband noises (0.2–6 kHz) with 10 ms on/off ramps presented at 70 dBA from speakers mounted in the horizontal plane ~1.2 m from the center of the listener's head. Spacing between speakers was 7.5°. The LEDs for the AV stimuli were mounted so that they were either horizontally aligned with the speakers or displaced (either to the left or to the right) by 5°. They were turned on and off in synchrony with the corresponding speakers. Two additional LEDs 10° below the speaker array served as fixation locations (azimuth~~s~~ of ±11.8°).

Procedures. Trials began with the onset of one of the two fixation LEDs. After subjects fixated the LED for 150 ms, the fixation LED was turned off and the AV or A-only stimulus was presented. The subjects performed a saccade to the perceived location of the stimulus. The saccade end point was recorded at the saccade end, i.e., when the eye fixation was sustained at the same location for 150 ms, at which point the experiment continued with the next trial. In both AV and A-only trials, the subjects were instructed to look to the location of the auditory component of the stimulus.

Training (AV) and probe (A-only) trials were randomly interleaved at a ratio of 1:1. Training stimuli were presented from one of the 3 training locations while the subject fixated the training fixation point (FP; top panel of Figure 1A). Probe stimuli were presented from one of the 9 speakers, while the subject fixated either the training or the non-training FP (bottom panel of Figure 1A).

Trials were run in sessions with a consistent AV pairing (leftward, rightward, or no shift). Each session started with a pre-adaptation reference measurement (18 A-only trials from the training fixation point), followed by 720 trials in which the training fixation point and the AV shift direction was fixed. Each subject performed 12 sessions (2 fixation points x 3 shift directions x 2 repeats) in order that was randomized across the subjects.

Data analysis. Data from the first quarter of each session were excluded to remove transitory values observed during the initial buildup of VAE. Within-session averages were computed from the remaining data separately for each combination of target location, fixation position, and condition. Since no large left-right differences were observed, data with training FP on the left were mirror-flipped and combined with the data with training FP on the right. All data are presented as across-subject means and standard errors of the mean.

Table 1: Four-way repeated-measures ANOVA of the VAE magnitude data

Factor	d.f.	F	Signif.
Speaker Location (1 to 9)	8, 48	33.87	***
A-only Fixation Point (Tr. vs. Non-Tr.)	1, 6	0.99	
Direction of Induced Shift (L vs. R)	1, 6	0.43	
AV Fixation Point (L vs. R)	1, 6	0.27	
Speaker Location X A-only FP	8, 48	0.79	*
Speaker Location X AV FP	4, 48	2.28	
A-only Fixation Point X AV FP	1, 6	0.42	
Speaker Location X Direction	8, 48	0.56	
AV Fixation Point X Direction	1, 6	2.16	
A Fixation Point X Direction	1, 6	0.1	
Speaker Loc. X AV FP X A-only FP	8, 48	0.31	
Speaker Loc. X AV FP X Direction	8, 48	0.52	
Speaker Loc. X A-only FP X Direction	8, 48	1.69	
AV FP X A-only FP X Direction	1, 6	0.12	
Loc. X AV FP X A-only FP X Direct.	8, 48	1.16	

Significance levels are as follows: * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.005$.

3.4 Overall Design and Results

As in Kopco et al. (2009), we presented paired visual-auditory stimuli in a subregion of audiovisual space, fixed in both eye- and head-centered coordinates. We used one initial eye fixation position on training trials and presented the discrepant audiovisual stimuli from a restricted spatial range that was lateral with respect to the fixation point (see Figure 1A, top). Because the visual training was local, we could test the spatial attributes of the resulting recalibration by shifting fixation on probe trials. Specifically, on interleaved auditory-only probe trials, we varied initial eye position (FP) with respect to the head (which was fixed) and presented sounds from all target locations spanning both the same head-centered locations and the same eye-centered locations as on the training trials (see Figure 1A, bottom). We first consider the effects observed on the AV training trials themselves before turning to aspects of how the effects generalize to the auditory-only conditions across both the trained and untrained regions of space as a function of eye-referenced vs. head-referenced fixation position.

3.4.1 Ventriloquism effect

A strong ventriloquism *effect* – or capture of the auditory stimulus location by the visual stimulus on combined AV trials - was observed. Green symbols in Figure 1B show the raw responses. When the AV stimuli were aligned, the average responses were not biased at all. The relative strength of the ventriloquism effect was evaluated as percent of shift in responses towards the visual (V) component re. the A-component on misaligned AV trials, which was for each A target location and V-component shift computed as $(resp_{V\text{-}misalign} - resp_{V\text{-}align}) / (stim_{V\text{-}misalign} - stim_{V\text{-}align})$, where *stim* is the actual location of the V-component. The strength ranged from 96% for the target at 15° to 82% for the target at 30° (averaged across 2 directions of induced shift). Even though there was a slight decrease in the strength of the ventriloquism effect for the most lateral targets, it was expected that, as in Kopco *et al.* (2009), this strong ventriloquism *effect* would be associated with a clear local ventriloquism *after*effect.

3.4.2 Gaze-dependent Effects During AV-aligned Baseline

We next assessed the auditory-only responses interleaved with the spatially aligned AV stimuli. The red and blue circles in Figure 1B show these responses. Overall, the pattern of results shows that the subjects accurately localized the auditory targets, showing a systematic displacement of the responses with the actual target locations. To analyze the impact of the visual training in more detail, the top panel of Figure 2A shows the biases in these responses relative to the actual target location, separately for the two fixation points. A gaze-direction-dependent adaptation is seen when comparing the responses from the training FP (red) to those from the non-training FP (blue). Specifically, the responses to the targets at approximately 10° azimuth were biased to the right by 2-3° when performed from the training FP (red “+” symbol) compared to the responses from the non-training FP (blue “+” symbol). A dashed line in this panel represents the same data from the central-adaptation experiment of Kopco *et al.* (2009), averaged across the two FP locations as no large FP-dependent differences were observed in that study. A solid black line in the bottom panel of Figure 2A shows the difference between the red and blue lines from the top panel, while the dashed line represents the difference from the central-adaptation experiment of Kopco *et al.* (2009). These panels show that responses to auditory-only stimuli from AV-trained locations that are lateral and near the training FP differ depending on whether eyes fixate within the same hemifield or the opposite hemifield. On the other hand, when the AV training locations are in the center, covering both hemifields, no such differential effect of fixation location is observed (dash-dotted line). ANOVA performed on the difference data showed a significant effect of target location ($F_{8,48} = 9.45$, $p < 0.00001$). This effect of eye fixation direction is strong, of size comparable to the VAE (see next section); thus, there is some eye-gaze-dependent contribution to responses to auditory-only stimuli even when vision is not used to induce any recalibration of the auditory spatial representation. However, this contribution is only visible if the AV stimuli are presented within one spatial hemifield. Overall, the pattern of results in the top panel of Figure 2A for both experiments is that, independent of FP location, the responses are very accurate in the trained region, while they tend to be biased away from the training region outside of it. This bias away is observed in all the non-training subregions for both FPs and both experiments, with the exception of the trained-FP data in the central region in the current experiment (3 red central targets in Figure 2A are approximately at 0°). Thus, the gaze-specific adaptation, which is observed in the same region, is likely caused by this lack of repulsion in the trained-FP central data.

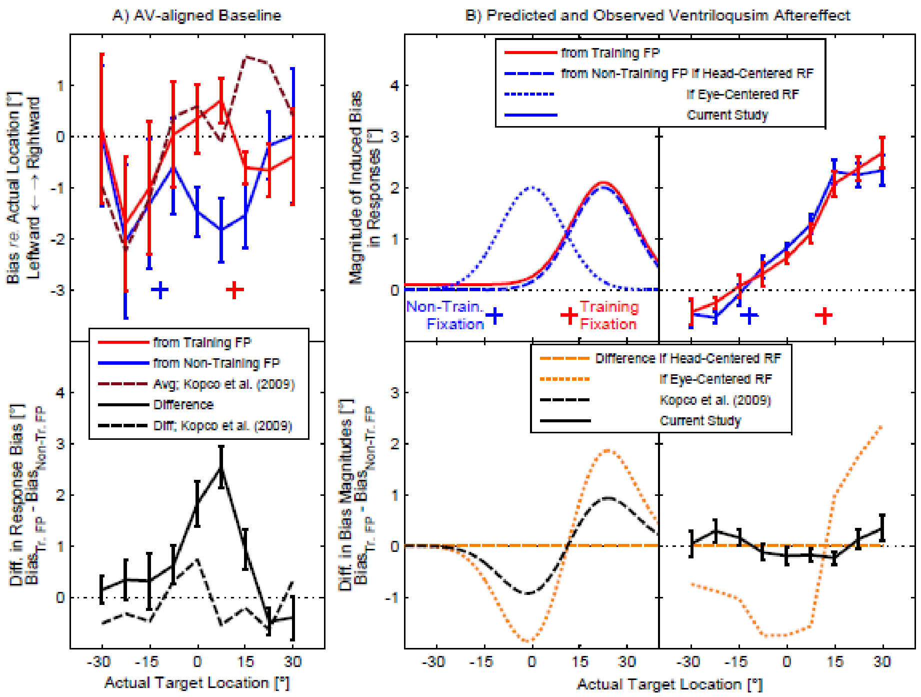

Figure 2: Adaptation induced by AV stimuli. A) Average bias in A-only responses in the AV-aligned baseline condition as a function of the actual target location. Top panel shows mean response biases (±SEM) when eyes are fixated at the training FP (red) and the non-training FP (blue). In addition, the across-FP average data for central adaptation from Kopco et al. (2009) are shown for comparison purposes (dashed line). The solid line in the bottom panel shows the difference between responses from training FP and the non-training FP. The dashed line shows the difference for taken from Kopco et al. (2009). B) Predicted and observed ventriloquism aftereffect. The top left panel plots the expected pattern of biases induced in the A-only probe responses when preceding AV trials are presented in the training region (15° - 30°). Red line shows predictions when the eyes fixate the training FP (i.e., the FP location used during AV training trials). Dash-dotted blue line shows expected results from the non-training FP if the RF of adaptation is head-centered, while dashed blue line shows expected results for an eye-centered RF. The bottom panel shows the differences between the expected bias magnitudes from the training versus the non-training FPs in the two RFs in orange. For comparison, the black dashed line sketches the results corresponding to the mixed RF observed after VAE was induced in the central region in (Kopco *et al.*, 2009) . Top right panel shows the across-subject mean (±SEM) difference between the auditory saccade end point locations when interleaved with spatially displaced AV stimuli vs. when interleaved with AV-aligned stimuli, collapsed across the direction of the AV displacement. The solid black line in the bottom right panel shows the effect of initial fixation position on the magnitude of the induced shift as the across-subject mean (±SEM) difference between the shifts from the training and non-training FPs (i.e the difference between the red and blue lines). Orange lines show the predictions of the difference for the two reference frames based on the training FP data (red) from the top right panel.

3.4.3 Ventriloquism Aftereffect and its Reference Frame

The expected pattern of ventriloquism aftereffect, and the predictions about the reference frame based on it, are illustrated in the left-hand panels of Figure 2B. The red line in the top left panel shows the predicted magnitude of the aftereffect induced by the AV stimuli, peaking in the trained region (15° - 30°) when assessed with eyes fixating the training FP. If visually induced spatial plasticity occurs in a brain area using a head-centered RF, then shifts in perceived sound location should occur mainly for sounds at the same head-centered locations (in Figure 2A, dash-dotted blue line matches the red line). Conversely, if plasticity occurs in an eye-centered RF, then visually induced shifts should occur mainly for sounds at the same eye-referenced locations (dotted blue line is shifted to the left of the red line by the same displacement as the non-training FP is shifted relative to the training FP). The bottom left panel summarizes the predicted results if evaluated as a difference between the responses from the training and non-training FPs. The dash-dotted orange line shows the difference between the red and dash-dotted blue lines, corresponding to the expected results if the reference frame is head-centered. The dashed orange line shows the difference between the red and dashed blue lines, corresponding to the expected results if the reference frame is eye-centered. The dashed black line shows the predicted difference in the biases expected if the RF is mixed, as observed in Kopco *et al.* (2009), in which case it should fall approximately in the middle of the predictions of the two RFs shown in orange.

We assessed the auditory-only responses interleaved with the spatially mis-aligned AV stimuli against these predictions. The red and blue triangles in Figure 1B show the raw responses in the conditions in which the ventriloquism aftereffect was induced in a leftward direction (leftward-pointing triangles) or rightward direction (rightward-pointing triangles). Overall, exposure to spatially mismatched AV stimuli resulted in a shift of responses to sounds in the direction of the previously presented visual stimuli (compare the corresponding triangles to the respective circles). To allow a detailed analysis of the results comparable with the predictions of Figure 2B, the red line in the top right panel of Figure 2B plots the magnitude of the bias in responses measured with eyes fixating the trained FP (red plus sign) *re.* no-shift baseline from Figure 1B, as a function of target location and averaged across the two directions of induced shift (note that no main effect or interaction involving the direction factor were significant in the ANOVA analysis, supporting this way of collapsing the data for visualization; Table 1).

The effect was strongest for the three right-most targets, i.e., in the trained region, reaching approximately 2.3° (51% of the ventriloquism effect strength). It was also location-specific, decreasing quickly toward zero outside of the trained region. These results are consistent with the results of Kopco *et al.* (2009), confirming that the VAE can be induced locally, so that it can be used to assess the VAE RF.

The reference frame of the VAE was examined by shifting the initial FP to a new location and examining how the observed VAE changed. The blue line in the top right panel plots the bias in responses measured with eyes fixating the new, non-trained FP (blue plus sign), shifted by approximately 23° to the left from the trained FP. There was very little difference in the measured VAE for the two FPs (blue line lies approximately on top of the red line). Thus, the observed results are consistent with visual–auditory recalibration occurring in a predominantly head-centered coordinate frame.

To compare the current results more directly to the predictions of the two models and to the data of Kopco *et al.* (2009), a difference between the shift magnitudes from the two FPs was computed (bottom right of Figure 2B, black traces) and compared with predictions based on the two models (orange traces). Again, the results are very close to the predictions of the head-centered RF.

These results were confirmed by performing a 4-way repeated-measures ANOVA with the factors of target speaker location (nine levels), fixation point of the trials (training vs. non-training FP), AV-trial fixation point location (left vs. right), and the direction of induced shift (left vs. right). The results of this analysis, summarized in Table 1, show that the main effect of location was always significant, confirming that the ventriloquism aftereffect is spatially specific and does not automatically generalize to the whole audiovisual field. The location by FP interaction was also significant, showing that the reference frame of visual–auditory recalibration is not purely head-centered, even though the eye-centered modulation is relatively small.

3.5 Discussion and Conclusions

The current study examined the spatial properties of the ventriloquism aftereffect induced by AV stimuli presented in only one spatial hemifield in the peripheral audio-visual field. The goal was to ascertain how the ventriloquism aftereffect unfolds as a function of multiple different spatial attributes: fixation position, generalization in head- vs. eye-centered coordinates, and training within one spatial hemifield in contrast to training in both hemifields (as in Kopco et al., 2009). The results indicate that the ventriloquism aftereffect is a multifaceted process, dependent on both the format of the neural representation of space in hearing and vision, and on the reference frame used by the two senses.

In terms of the representational format, the location of the fixation position impacted the pattern of adaptation induced by the AV stimuli, even when the AV-stimuli were presented from matching locations and no VAE was induced. This unexpected adaptation was not observed in the previous central-adaptation study (Kopco *et al.*, 2009). And, it is difficult to identify its cause, since a baseline measurement with no AV stimulation was not performed. However, a comparison of the central-adaptation and peripheral-adaptation data suggests that adaptation away from the training region was observed in the AV-aligned data in both experiments. Such expansion of space is consistent with previously observed inherent biases towards the periphery (Razavi *et al.*, 2007). The current data shows that the inherent biases might be more correctly described as biases away from the AV-training region, rather than towards the periphery, and that the biases might be modulated by eye-gaze direction. Specifically, in the current experiment in which the AV-aligned stimuli were presented in the periphery, there was no repulsive bias in the central region when the gaze was fixated to a point in the training hemifield, but it was observed if the gaze was fixated in the opposite hemifield. At least two other factors of the current experimental design might also contribute to the effect. First, the effect might be a result of adaptation to the auditory stimulus-distribution, which becomes skewed when the training stimuli are included since all of them come from one side (e.g., similar to adaptation reported by Dahmen *et al.* (2010). Second, the visual signal might be causing some global ventriloquism-like adaptation outside the training region, such that the auditory-only responses are shifted towards the region from which the visual stimuli are frequently presented, but only when the FP is in the hemifield ipsilateral to the AV stimulation (and such shift towards the training region cancels out the repulsion observed otherwise). Whatever the specific mechanism, this adaptation effect shows that there is a hemifield-specific integration of visual and auditory spatial signals that differs from the integration occurring when the stimuli are presented centrally, covering both spatial hemifields.

Regarding reference frames, the current results together with those of Kopco et al. (2009) show that in humans the RF of VAE is a mixture of eye-centered and head-centered coding. In the central region, the effect is a fairly even mixture of these two reference frames, whereas in the periphery, the pattern more closely fits the head-centered predictions, but also shows an interaction with eye position. This shows that the transformation of the visual and auditory signals into an aligned reference frame, thought to be necessary for the ventriloquism aftereffect to work, is non-uniform. While it is not immediately clear what form of non-uniformity might be causing this pattern of results, it may be related to the hemispheric-difference channel models of auditory space representation (Salminen *et al.*, 2009; Grothe *et al.*, 2010; Groh, 2014).

Note that we also performed these experiments in two rhesus monkeys, as was the case in Kopco et al. (2009) (footnote 1). In brief, in the monkeys there was no evident difference between the RF pattern observed in the central and peripheral region, which always mixed between head- and eye-centered frames, consistent with most neurophysiological observations in the same species (Moriya *et al.*, 2013). Overall, these differences across training regions (and, possibly, across species) suggest that the locations in the brain that are recruited to accomplish this recalibration of auditory space may be widely varied. Some are likely head-centered, some are eye-centered, some may involve the position of the eyes in the orbits per se. These sites of plasticity may be recruited differently depending on the training region and whether it spans both head-centered hemifields or is contained within one.

Additional experimental and/or modeling studies are needed to test alternative explanations about the different reference frames of the ventriloquism aftereffect as well as about the unexpected AV-aligned adaptation effect. However, the current results demonstrate that there are hemisphere-specific adaptation processes in visual recalibration of auditory space, resulting in different FP-dependent patterns of adaptation depending on the region in which adaptation is induced.

Acknowledgments

This work was supported by the SRDA, project DS-2016-0026, EU H2020-MSCA-RISE-2015 grant no. 69122, and by the EU RDP projects TECHNICOM I, ITMS: 26220220182, and TECHNICOM II, ITMS2014+:313011D23. Barbara Shinn-Cunningham was supported by NIH R01 DC013825 Jennifer Groh was supported by NIH NS50942 and NSF 0415634.

4 A preliminary model of the reference frame of the ventriloquism aftereffect I.

The current chapter contains the content of the article Loksa and Kopco (2016). All work described here was performed by the thesis author under the supervision of the thesis advisor.

4.1 Abstract

The human brain extracts information from various senses in order to represent the physical space. To integrate spatial information from the visual and auditory modalities, the modalities need to be aligned as each of them represents spatial information in a different reference frame (RF). The visual reference frame is aligned with the direction of eye gaze while the auditory one is aligned with the orientation of the head. The aligned audiovisual spatial representation is most likely using one of the two RFs as well. Previous experimental data attempting to identify the aligned RF are inconsistent. This article presents modeling attempts aiming at resolving this inconsistency and identifying the reference frame of the ventriloquism aftereffect.

4.2 Introduction

Vision plays an important role in how the brain processes auditory information (Alais and Burr, 2004). In the spatial domain, vision provides guiding signals for calibration of spatial auditory processing. This can be illustrated by the ventriloquism aftereffect illusion in which repeated pairings of spatially mismatched visual and auditory stimuli produce shifts in the perceived locations of sound sources that persist even when the sounds are presented by themselves (Knudsen and Knudsen, 1985; 1989; Alais and Burr, 2004). It might be that the supramodal spatial representation could be the ultimate one in sense of being directly used in motion planning etc.

The current study models data from a previous study which examined the RF of the ventriloquism aftereffect (abbr.: RFVAE) (Kopco *et al.*, 2009). RFVAE might by identical or connected to RF of general supramodal spatial adaptation.

There were two basic hypotheses considering properties of RFVAE so would be: (1) head- and (2) eye-centered, in case of holding of which the RF is spatially fixated to specific body part (1) head itself (2) eyeball. The reason for choosing such ones as possible RF-s is because respectively (1) auditory and (2) visual localization go on in these RF-s (Brainard and Knudsen, 1995; Razavi *et al.*, 2007).

In a previous study attempting to identify RFVAE the so called aftereffect magnitude was compared between following two conditions: eye not shifted from position of ventriloquism aftereffect inducement, eye shifted so. By eye shift we mean change of so called initial fixation point, in which the eye is right before providing the stimulus. And if such aftereffect magnitude shifted with eye, RFVAE would be probably eye-centered. If it didn't shift, it would be head centered, since head shifts neither.

Modelling was performed because of inconsistency of results according to basic hypotheses. In current article we will also show such basic results first.

4.3 Experimental data

The experimental data used here are taken from a previous study that investigated the reference frame of ventriloquism aftereffect (Kopco *et al.*, 2009).

4.3.1 Materials and methods

Figure 3 illustrates the experimental setup and the hypothesized results.

In experiment given subject was sitting in dark quiet room with his head fixed. The target speakers and LEDs (visual adaptor) were placed in order to provide stimuli to subject. The saccadic responses to stimuli were recorded.

To induce ventriloquism aftereffect the AV training trials with constant shift of light from sound were induced in specific azimuth region, while FP-s of all such trials are same within session (training fixation point (TrFP)).

In order to measure aftereffect magnitude in condition of eye not shifted from position of ventriloquism aftereffect inducement, the localization errors were identified according to responses to auditory-only (A-only) trials in TrFP in stimuli range -30° to 30°. Analogically was done for condition of eye shifted in so called Non-training fixation point (NTrFP). So within session AV trials were in TrFP and there were A-only trials in TrFP and A-only trials in NTrFP. These three kinds of trials were interleaved.

To see whether ventriloquism aftereffect is symmetrical or not, the session differed in (1) in shift of visual component of AV trials from its auditory component, and (2) in training fixation point. There were three kinds of shifts of visual component: no shift (sound and light have same azimuth), positive shift (visual component is shifted by 5° to the side, toward which the TrFP is from 0°). Regarding FP-s azimuth axis can be flipped that TrFP would on 11.8° and NTrFP on -11.8° for each session.

Because discriminability in center vs. periphery is inconsistent (Maier *et al.*, 2010), two different so called training regions of aftereffect inducement were used, but the same one within each session. These two we call center and periphery. In Figure 3 central one is marked.

The 9 speakers were displaced within same horizontal plane, while holding: distance of each speaker from center of the listener's head is equal; angle difference of the speaker from adjacent one is equal (7.5°) (see Figure 3).

According to Diff in bias magnitude (bias of NTrFP A-only trials subtracted from bias of TrFP A-only trials; Figure 3) the RFVAE had to be identified.

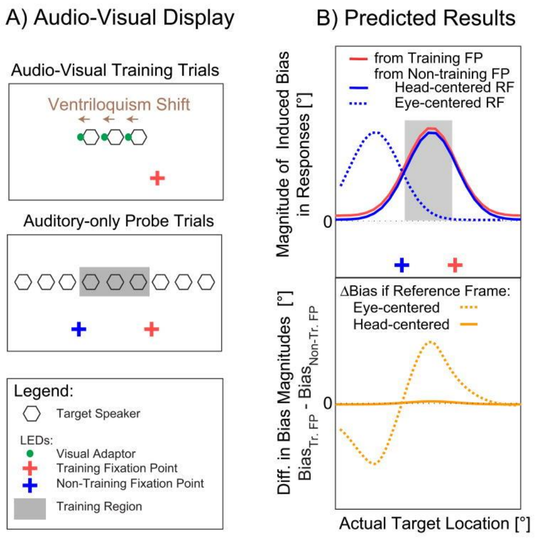
A) Audio-Visual Display
Audio-Visual Training Trials
Ventriloquism Shift
Auditory-only Probe Trials
Legend:
Target Speaker
LEDs:
Visual Adaptor
Training Fixation Point
Non-Training Fixation Point
Training Region
B) Predicted Results
from Training FP
from Non-training FP
Head-centered RF
Eye-centered RF
Magnitude of Induced Bias in Responses [°]
0
ΔBias if Reference Frame:
Eye-centered
Head-centered
Diff. in Bias Magnitudes [°]
Bias$_{Tr. FP}$ - Bias$_{Non-Tr. FP}$
0
Actual Target Location [°]

Figure 3: A) Symbols in "Audio-Visual Training Trials" panel mark the azimuths of stimuli provided to subjects in audiovisual training trials, in the way that the azimuthal relative shift between physical location of stimuli, that are synchronous, are constant within given experimental session for each session. The symbols in "Auditory-only Probe Trials" panel mark azimuths of auditory-only trials, which were interleaved with the already mentioned training ones. B) This panel visualize hypothetical experimental data for cases of questioned reference frame being head- vs. eye centered. "Magnitude of Induced Bias in Responses" here means localization error of them toward the shift in given session for each session.

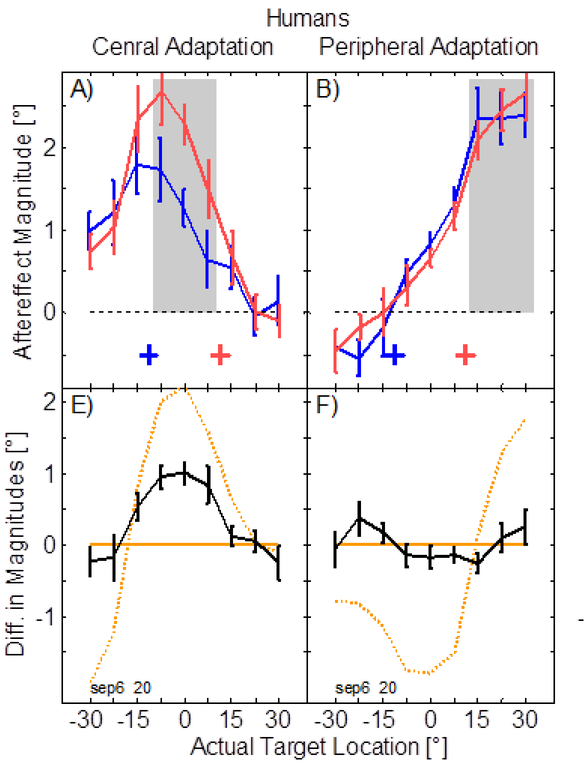

Figure 4: Magnitude of ventriloquism aftereffect and reference frame determination according to difference between training vs. non-training trials. Red/Blue line - separation of probe auditory-only trials according to the pre-trial eye gaze azimuths (marked by '+' of given color). But eye gazes of all audio-visual training trials are preceded by the red one so this is called training fixation point (FP), and the blue one non-training one. The black line can be arithmetically described as the subtraction of blue line from the red line and we call it aftereffect FP dependence. The orange lines reflect hypothetical Aftereffect FP dependences: the solid one for the case of eye-centered head centered and the dotted one.

4.3.2 Early analysis

According to Figure 4 the result of RFVAE is inconsistent: according to central adaptation this RF seems to be mixed of head- and eye centered, while for peripheral adaptation it seems to be purely head-centered.

In order to resolve this inconsistency, we attempted to model experimental data. But in order to model it we first displayed results for different conditions Figure 5.

It was unlikely in the brain that two different forms of reference would be utilized for same representation. On the other side, there are multiple other explanations for difference we observed, related to other forms of adaptation that might have occurred in this experiment: the undershooting, the expansion of auditory space, saccade adaptation. The Kopco *et al.* (2009) data showed another form of plasticity for which the study has not been designed, and in modelling we will explore first two above-mentioned explanations.

Undershooting means shortening saccades in comparison of them in case when saccade endpoints ended in location where the stimulus is perceived.

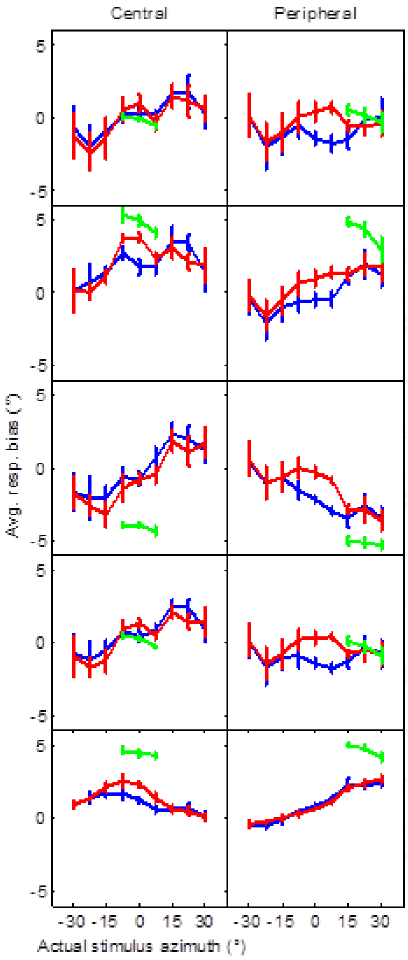

Figure 5: Mean localization error of human subject experimental data and SEM across 7 subjects. Red line – A-only trials - training fixation point, blue line A-only data – non-training fixation point, green line – AV (training) trials, black line – difference between training vs. non-training A-only trial mean (FP dependence), magenta line – difference between peripheral vs. central adaptation FP dependence. Conditions according to rows respectively: 1. no shift, 2. positive shift, 3. negative shift, 4. mean across shifts, 5. aftereffect magnitude. The graphs in the 5th except of magenta lines row are little different with 0 A, B, E, F, except of yellow lines only because of technical errors and outliers removal.

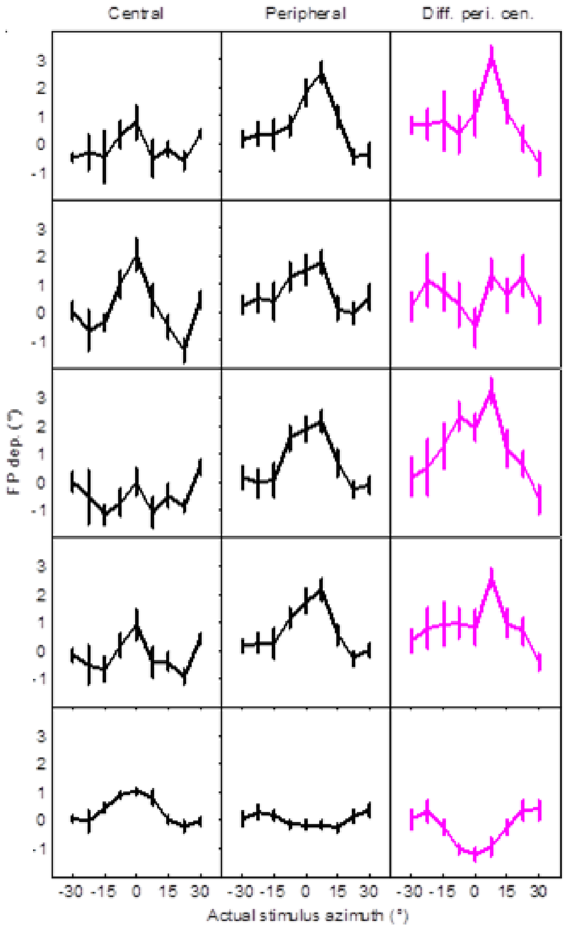

Figure 6: Continuation of Figure 5.

4.4 Unexpected form of plasticity

In Figure 7 we observed inconsistency. In this figure we can see different azimuth and different condition that there are two types of cases for localization error being (1) depending (2) not depending on initial eye fixation point visualized as (1) similar or (2) dissimilar value of red vs. blue line: 1. all central azimuths, azimuths -30° to -15° in periphery and azimuths 15° to 30° in periphery. 2. azimuths -7.5° to 7.5° in periphery. This unexpected plasticity could be possible reason for inconsistency of central vs. peripheral RF-s of ventriloquism aftereffect appearance.

In order to explain this we attempted to model data present in this visualization (Typical property of this visualization is consistency of audiovisual training trials that affect localization errors (so called no-shift) as the selection key for data included.

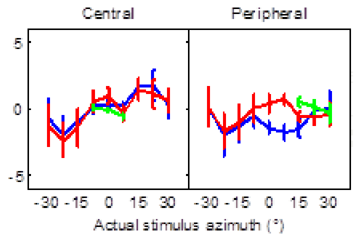

Figure 7: Localization error for no-shift condition for different training regions.

4.5 Modelling

In this section there is the attempt to model newly observed phenomenon, and also test of relevant qualities of its result using a model that assumes that two adaptive processes occur, unrelated to the ventriloquism aftereffect, and that their effect combines additively.

4.5.1 Description

Basic idea of this modelling is to consider following two factors in additive relation: saccade hypometry, expansion outside training region.

Saccade hypometry is in other words undershooting of saccades (Harris, 1995). Saccade hypometry shortens the saccade in comparison with the case when saccade would end in location where the response is perceived. We considered it because according to Figure 7 peripheral data, azimuths -7.5° to 7.5° the localization errors appear to be shifted from each other toward related fixation point.

The effect which we call outside training region expansion has zero value inside training region and its absolute value increases with distance from this region. The reason for this is that the data appear to reflect this phenomenon.

Established variables and functions:

t ...target azimuth

$b(t, FP, trreg)$ ...bias of response from auditory stimulus location to target at azimuth t from eyes initially fixated at FP when the training region was $trreg$ (predicted variable).

$eotr(t, trreg)$ ...expansion of response on target with azimuth t outside training region $trreg$

$h(t, FP)$ ...saccade hypometry on target t with eyes initially fixated on FP

$dtr(t, trreg)$ ...relative distance of target t from training region $trreg$

Free parameters: $ak_1, ak_2, ek_1, \; ek_2$

Established equations:

$b(t, FP, trreg) = eotr(t, trreg) + h(t, FP);$ ... additive relation of given concepts.

$$eotr(t, trreg) =$$

$$= ek_1 \cdot \left(\frac{1}{1 + e^{ek_2 \cdot (dtr(t, trreg))}} - \frac{1}{2}\right);$$

$$h(t, FP) =$$

$$= ak_1 \cdot \left(\frac{1}{1 + e^{ak_2 \cdot (t - FP)}} - \frac{1}{2}\right);$$

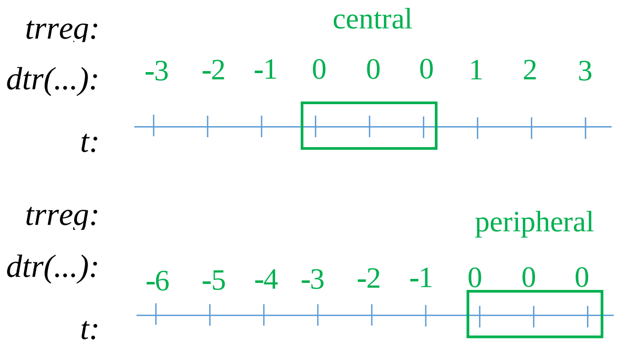

Figure 8: Illustration of the relative distance of t from training region $trreg$. $trreg$ is training region, t is target azimuth and dtr is relative distance of t from training region $trreg$. (There is arithmetical distinction between this and the addend of the model, but that depend on this, according to equation in current subchapter)

4.6 Performance

The free parameters for model were fitted in MATLAB function nlinfit was used which is using iterative least squares estimation. The data in Figure 7 are used as the base for observational data for this model.

In first step the hypometry was fitted. This was done in domain of FP dependence. This means that observational data and fitted model as the input to this tool were converted to this format. Results are displayed in Figure 9**Chyba! Nenašiel sa žiaden zdroj odkazov.**. In this step parameters ak_1 and ak_2 were fitted.

In second step was done on residual of data after subtracting results of the first step, expansion outside training region was fitted. This was done in domain of bias (localization error). Results are displayed in 0 In this step parameters ek_1 and ek_2 were fitted.

Resulting coefficients are following:

$ak_1 = 0.94,$
$ak_2 = 151.13$
$ek_1 = -2.26$
$ek_2 = 130.62$

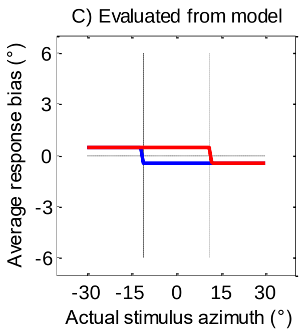

Figure 9: The visualization of the saccade hypometry according (h) to the resultant model fitted on experimental data. Colors have meaning analogical to whole article.

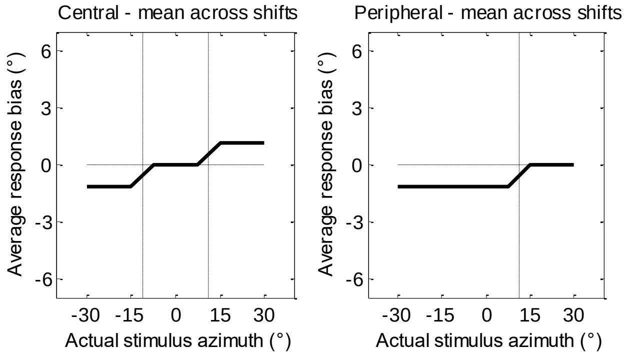

Figure 10: The visualization of the expansion outside training region according ($eotr$) to the resultant model fitted on experimental data.

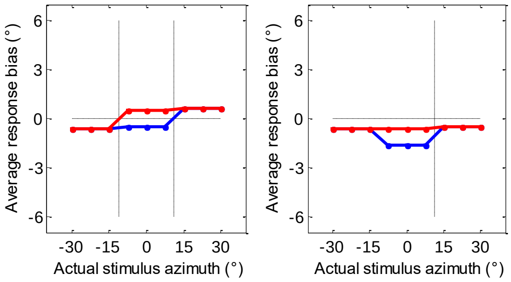

Figure 11: approximation of behavior of model of bias (b) in no-shift condition for central (left graph) and peripheral (right graph) adaptation according to given model. Colors have meaning analogical to whole article.

Figure of current modelling results (Figure 11) show no difference between FP dependences of central vs. peripheral adaptation (FP dependence is difference between training vs. non-training fixation biases. (Difference in red vs. blue)). Absence of such difference is inconsistent with experimental data (Figure 7) and with explanation of unexpected form of plasticity.

In Figure 11 there were some FP independent conditions (-30° to -15° and 15° to 30° for both training regions) and some FP dependent conditions (-7.5° to 7.5° for both training regions). There is the difference with experimental data because for -7.5° to 7.5° for model data they are independent.

4.6.1 Proof of current model inappropriateness

We can even prove mathematically that current model cannot explain current unexpected form of plasticity.

To show this, we define:

$FPdep(t, trreg)$ ...FP dependence of biases on training region $trreg$

$TrFP$...azimuth of training fixation point.

$NtrFP$...azimuth of non-training fixation point

$$FPdep(t, trreg) =$$
$$= b(t, TrFP, trreg) - b(t, NtFP, trreg);$$

We infer:

$$FPdep(t, trreg) =$$
$$= b(t, TrFP, trreg) - b(t, NtrFP, trreg) =$$
$$= eotr(t, trreg) + h(t, TrFP) -$$
$$(eotr(t, trreg) + h(t, NtrFP)) =$$
$$h(t, TrFP) - h(t, NtrFP);$$

We see that according to current model FP dependence ($FPdep$) does not depend on training region ($trreg$), so the model cannot describe such a dependence.

In experimental data FP dependence depend on training region in azimuth range -7.5° to 7.5°. But according to this proof model cannot do so.

4.7 Conclusion

We have described a previous study examining the reference frame of the ventriloquism aftereffect and its main results, which contain some ambiguity. We examined a part of the experimental data from that study, and we described a new adaptive phenomenon. We made the attempt to model these data and we have proven that the proposed additive model combining hypometry and auditory space expansion is inappropriate for the explanation of the newly observed phenomenon in manner.

One of the alternatives for current modelling is instead of additivity of factors the function composite would be used, so these factors would be related hierarchically such that one operates on the output of the other one. Alternatively, completely other factors might play role. Additional modeling is currently underway to examine these alternatives.

4.8 Acknowledgement

This contribution was supported by Slovak Research and Development Agency under grant project APVV-0452-12 Spatial Attention and Listening in Complex Environments.

5 A preliminary model of the reference frame of the ventriloquism aftereffect II.

The current chapter contains the content of the article Loksa and Kopco (2017). All work described here was performed by the thesis author under the supervision of the thesis advisor.

5.1 Abstract

The reference frame (RF) used by audio-visual (AV) spatial representation is likely to be head-centered or eye-centered, aligned with the RFs of either the unimodal auditory (head-centered) or visual (eye-centered) representations. Results of previous RFAV studies are inconsistent, suggesting that the RF is either mostly head-centered, when examined in the periphery, or a mixture of head-centered and eye-centered, when examined in the central field (Kopco *et al.*, 2009; Loksa and Kopco, 2016). Here, a model is proposed, assuming a form of a priori bias is combined with the adaptation due to AV stimuli. This model can explain the results in the baseline conditions, but not when ventriloquism aftereffect is induced. Therefore, additional mechanisms are likely to determine the AV RF.

5.2 Introduction

Vision plays an important role in how the brain processes auditory information (Alais and Burr, 2004). In the spatial domain, vision provides guiding signals for calibration of spatial auditory processing. This can be illustrated by the ventriloquism aftereffect illusion in which repeated pairings of spatially mismatched visual and auditory stimuli produce shifts in the perceived locations of sound sources that persist even when the sounds are presented by themselves (Knudsen and Knudsen, 1985; 1989; Alais and Burr, 2004). It might be that a supramodal spatial representation exists, directly used in motion planning etc.

The current study models data from a previous study which examined the RF of the ventriloquism aftereffect (RFVAE) (Kopco *et al.*, 2009). RFVAE might by identical or connected to RF of general supramodal spatial adaptation.

There were two basic hypotheses considering properties of RFVAE so would be: (1) head- and (2) eye-centered, in case of holding of which the RF is spatially fixated to specific body part (1) head itself (2) eyeball. The reason for choosing such ones as possible RFs is because respectively (1) auditory and (2) visual space is represented in these RFs (Brainard and Knudsen, 1995; Razavi *et al.*, 2007).

In a previous study of the RFVAE, the observed aftereffect was compared between two conditions: eyes not shifted from the fixation point (FP) of ventriloquism aftereffect inducement, eye shifted to a new FP. By eye shift we mean change of fixation position, i.e., direction of the eye gaze when the stimulus is presented. It was hypothesized that if the aftereffect shifted with the eye shift, RFVAE would be probably eye-centered. If it didn't shift, it would be head centered, since head shifts neither. The goal of the previous modelling was to evaluate a possible mechanism causing this inconsistency of results with respect to the above hypotheses.

In the current study we first show the behavioral results, followed by an extension of the model and its evaluation.

5.3 Experimental data

The experimental data used here are taken from a previous study that investigated the reference frame of ventriloquism aftereffect (Kopco *et al.*, 2009).

5.3.1 Materials and methods

Figure 12 illustrates the experimental setup and the hypothesized results. In the experiment the subject was sitting in a dark quiet room with his head fixed. The target speakers and LEDs (visual adaptor) were used to provide stimuli to the subject. The saccadic responses to stimuli were recorded.

To induce ventriloquism aftereffect the AV training trials with constant shift of light from sound were induced in specific azimuth region, while FPs of all such trials were same within session (training fixation point (TrFP); Figure 12A).

To measure the aftereffect magnitude in the condition of eye not shifted from position of ventriloquism aftereffect inducement, the localization errors were identified according to responses to auditory-only (A-only) trials in TrFP in stimuli range -30° to 30°. Analogically was done for condition of eye shifted in so called Non-training fixation point (NTrFP). So within session AV trials were in TrFP and there were A-only trials in TrFP and A-only trials in NTrFP. These three kinds of trials were interleaved.

To see whether ventriloquism aftereffect is symmetrical or not, the session differed in (1) in shift of visual component of AV trials from its auditory component, and in training fixation point. There were three kinds of shifts of visual component: no shift (sound and light have same azimuth), positive shift (visual component is shifted by 5° to the side, toward which the TrFP is from 0°), or negative shift (opposite of positive shift). Regarding FPs azimuth axis can be flipped that TrFP would on 11.8° and NTrFP on -11.8° for each session.

Because discrimination abilities in center vs. periphery are inconsistent (Maier et. al., 2009), two different training regions of aftereffect inducement were used, but the same one within session. These two we call center and periphery. In Figure 12 central one is shown.

The 9 speakers were displaced within same horizontal plane, while holding: distance of each speaker from center of the listener's head is equal; angle difference of the speaker from adjacent one is equal (7.5°) (see Figure 12).

According to diff. in bias magnitude (bias of NtrFP A-only trials subtracted from bias of TrFP A-only trials; Figure 12) the RFVAE had to be identified. Results are displayed in Figure 13.

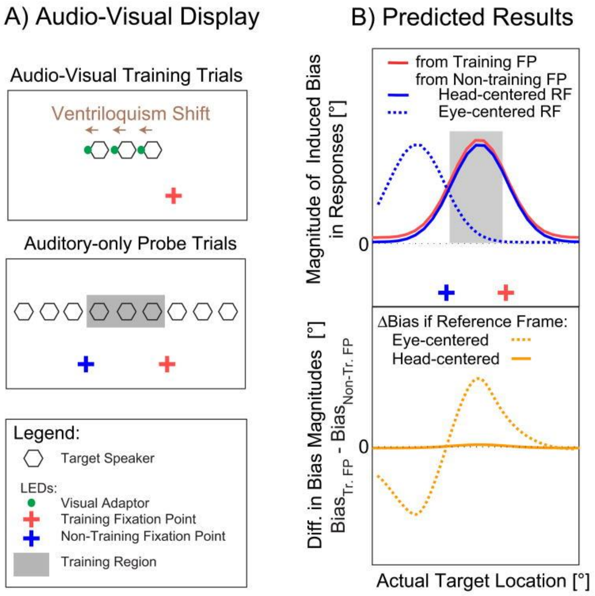
A) Audio-Visual Display
B) Predicted Results
Audio-Visual Training Trials
Ventriloquism Shift
Auditory-only Probe Trials
Legend:
Target Speaker
LEDs:
Visual Adaptor
Training Fixation Point
Non-Training Fixation Point
Training Region
from Training FP
from Non-training FP
Head-centered RF
Eye-centered RF
Magnitude of Induced Bias in Responses [°]
0
ΔBias if Reference Frame:
Eye-centered
Head-centered
Diff. in Bias Magnitudes [°]
$\mathrm{Bias}_{\mathrm{Tr.\ FP}}$ - $\mathrm{Bias}_{\mathrm{Non\text{-}Tr.\ FP}}$
0
Actual Target Location [°]

Figure 12: A) Symbols in "Audio-Visual Training Trials" panel mark the azimuths of stimuli provided to subjects in audiovisual training trials, in the way that the azimuthal relative shift between physical location of stimuli, that are synchronous, are constant within given experimental session for each session. The symbols in "Auditory-only Probe Trials" panel mark azimuths of auditory-only trials, which were interleaved with the already mentioned training ones. B) This panel visualize hypothetical experimental data for cases of questioned reference frame being head- vs. eye centered. "Magnitude of Induced Bias in Responses" here means localization error of them toward the shift in given session for each session.

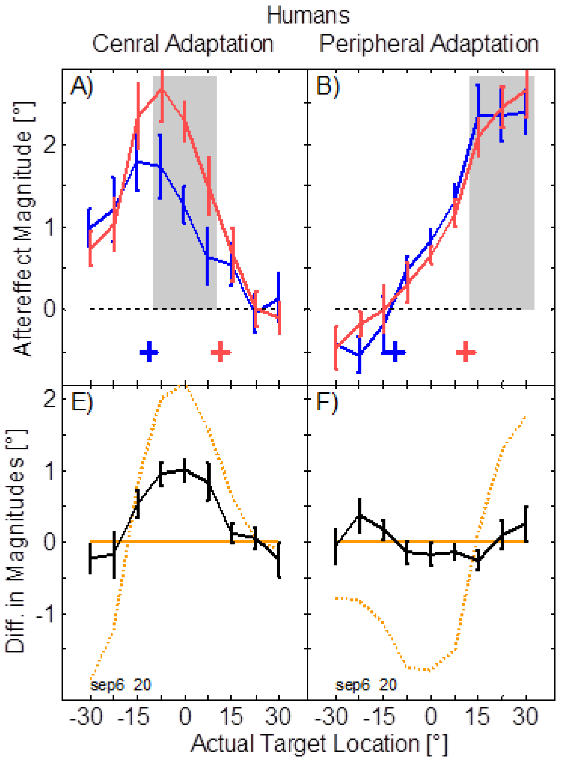

Figure 13: Magnitude of ventriloquism aftereffect and reference frame determination according to difference between training vs. non-training trials. Red/Blue line - separation of probe auditory-only trials according to the pre-trial eye gaze azimuths (marked by '+' of given color). But eye gazes of all audio-visual training trials are preceded by the red one so this is called training fixation point (FP), and the blue one non-training one. The black line can be arithmetically described as the subtraction of blue line from the red line and we call it aftereffect FP dependence. The orange lines reflect hypothetical Aftereffect FP dependences: the solid one for the case of eye-centered head centered and the dotted one.

5.3.2 Data analysis

Figure 12 shows that the result of RFVAE is inconsistent in the two regions: for central adaptation this RF seems to be mixed of head- and eye centered, while for peripheral adaptation it seems to be purely head-centered.

To resolve this inconsistency, we attempted to model the experimental data. To better understand the causes of the inconsistency, Figure 14 and Figure 15 show the detailed behavioral results for different conditions.

It is unlikely in the brain that two different forms of reference would be utilized for the same representation. On the other side, there are multiple other explanations for difference we observed, related to other forms of adaptation that might have occurred in this experiment: the saccadic hypometria (undershooting), the expansion of auditory space, saccade adaptation.

The Kopco *et al.* (2009) data showed another form of plasticity, described in the following section,for which the study has not been designed, and in modelling first two above-mentioned explanations were explored in previous article Loksa and Kopco (2016).

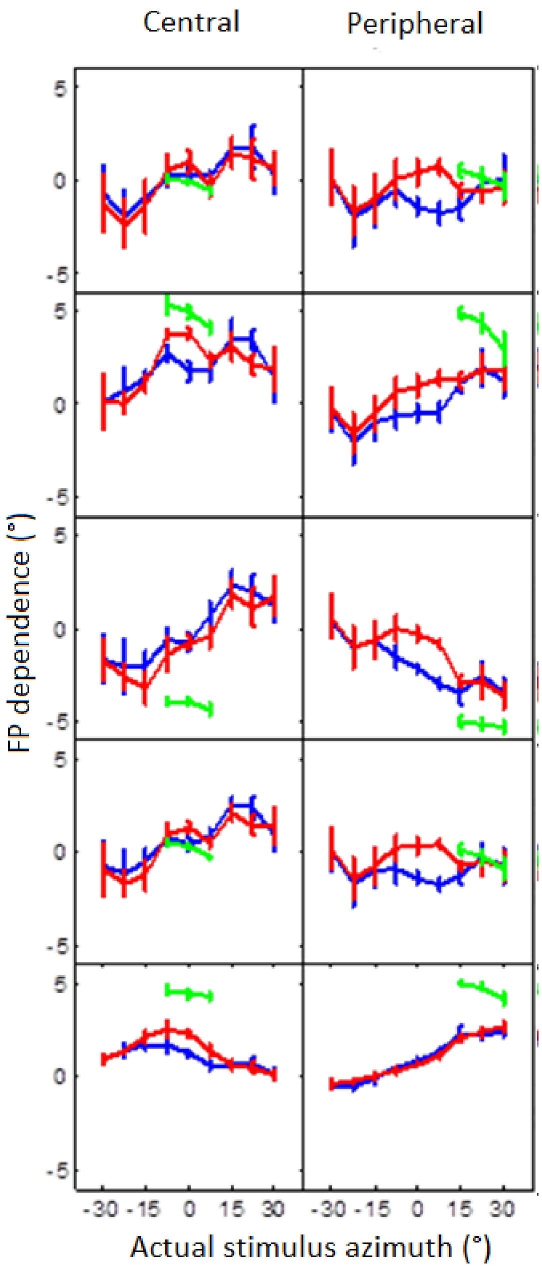

Figure 14: Mean localization error of human subject experimental data and SEM across 7 subjects. Red line – A-only trials - training fixation point, blue line A-only data – non-training fixation point, green line – AV (training) trials, black line – difference between training vs. non-training A-only trial mean (FP dependence), magenta line – difference between peripheral vs. central adaptation FP

dependence. Conditions according to rows respectively: 1. no shift, 2. positive shift, 3. negative shift, 4. mean across shifts, 5. aftereffect magnitude. The graphs in the 5th except of magenta lines row are little different with Figure 13: A, B, E, F, except of yellow lines only because of technical errors and outliers removal.

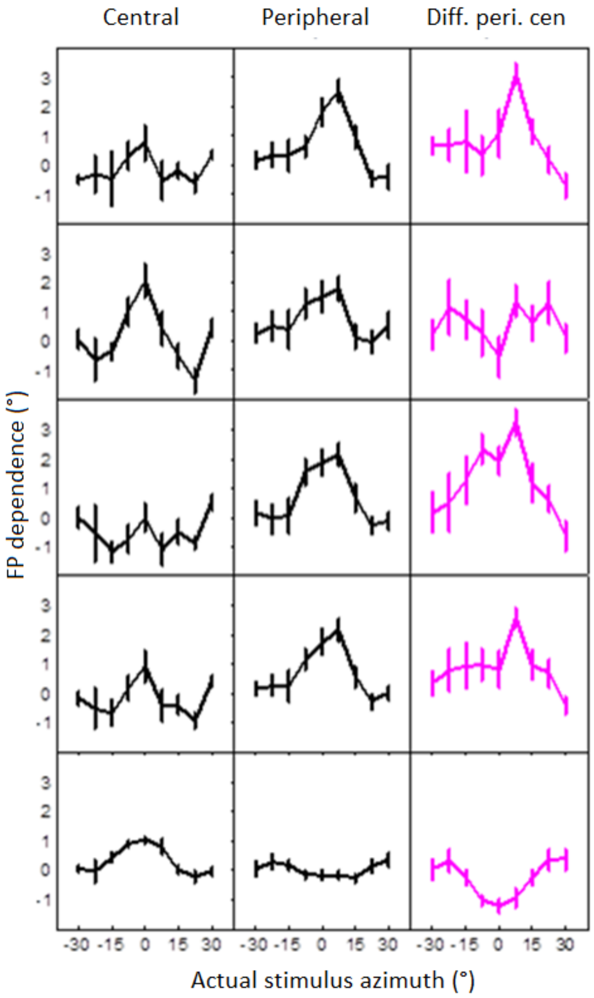

Figure 15: Continuation of previous figure.

5.4 Unexpected form of plasticity

In Figure 16 we observed inconsistency. In this figure we can see different azimuth and different condition that there are two types of cases for localization error being (1) depending (2) not depending on initial eye fixation point visualized as (1) similar or (2) dissimilar value of red vs. blue line: 1. all central azimuths, azimuths -30 to -15 in periphery and azimuths 15 to 30 in periphery. 2. azimuths -7.5 to 7.5 in periphery. This unexpected plasticity could be possible reason for inconsistency of central vs. peripheral RFs of ventriloquism aftereffect appearance.

In order to explain this inconsistency we attempted to model data present in this visualization (Typical property of this visualization is consistency of audiovisual training trials that affect localization errors (so called no-shift) as the selection key for data included.

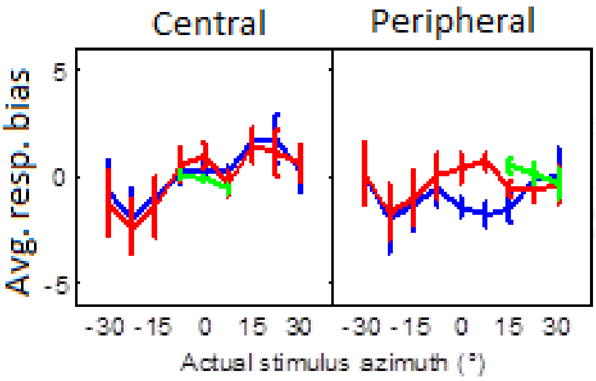

Figure 16: Localization error for no-shift condition for different training regions (variables are in °).

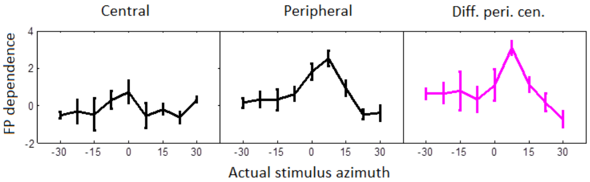

Figure 17: Continuation of previous figure. FP dep. means difference between biases of TrFP vs. NtrFP. (The magenta graph for model of (Loksa and Kopco, 2016) would be zero constant - black ones would be equal)

5.5 Modelling

This section presents a model of the newly observed adaptation, and also tests of relevant qualities of the model. The model assumes that a priori bias is modified by so called AV effect.

Previous modelling assumed that unexpected form of plasticity is caused by saccadic hypometry and expansion outside training region, specifically their additive composite, and this was proven to be insufficient to describe the data (Loksa and Kopco, 2016).

5.5.1 Description of the current model

Basic idea of this modelling is a priori bias affected by biases of responses to AV stimuli (vertical position of green line). A priori bias is attracted by vertical position of green dots. Attraction is represented by weighted mean of A priori bias and vertical positon of green points. Weight of a priori bias is one of the free parameters (w_p). Weight of given vertical position of green point is product of (1) implicit free parameter ($1-w_p$) and of (2) Gaussian function of: horizontal position of green point as the center of Gaussian function, azimuth of auditory-only stimulus as the main input of it and another free parameter as the width of the Gaussian function (wdt).

A priori bias is sigmoid function modified to be odd (inflection point at vertical 0 instead of 0.5), where its horizontal center ($-FP \cdot c$), its height (hg) and its slope (sp) are adjustable by free parameters.

5.5.1.1 Established variables and functions

Free parameters (Tab. 1:, parameters that were set by non-linear fitting algorithm of MATLAB (nlinfit)):

c	coefficient of horizontal position of inflection point according to fixation point for a priori bias.
hg, sp	height and slope of a priori bias.
w_p	weight for a priori bias.
w_{wdt}	width of AV effect.

Tab. 1: Free parameters.

Input variables (Tab. 2:, vector of these variables determine prediction case):

a_{A-st}	actual azimuth of stimulus.
a_{FP}	azimuth of fixation point (red vs. blue line).

a_{AV-st}	vector of azimuths of A component of AV stimuli (these are marked as value of horizontal coordinate of green error bar centers).
$b_{AV-resp}$	vector of biases of response to AV stimuli (these are marked as value of vertical coordinate of green error bar centers).

Tab. 2: Input variables.

Established equations:

A priori bias is represented (left side of equation) and defined (right side of equation) in following equation (It is sigmoid moved vertically in order for have value domain in $[-hg, hg]$)

$$b_p(x, a_{FP}) = 2 \cdot hg \cdot [1 + \mathrm{e}^{z}]^{-1} - 1;$$
$$z = \text{-}sp \cdot [x\text{-}(\text{-}a_{FP} \cdot c)];$$
$$c, hg, sp > 0;$$

We mark the Gaussian function as following

$Gauss(x, \mu, \sigma)$; where x is the main variable, μ is center and σ represents width of this function.

We mark and define the width function:

$$f_{wdt}(x, \mu, \sigma) = \frac{Gauss(x, \mu, \sigma)}{\sum_{i=-8}^{8} Gauss(7.5 \cdot i, 0, \sigma)} \; .$$

This function is the function of closeness of x to μ.

The main function is implemented below, and its result is bias of response to A-only stimuli and the model predicts this bias:

$$b_{A-resp}=\frac{F_1\cdot w_p+\sum_{i=1}^{count_{AV}}F_2(i)\cdot w_2(i)}{w_p+\sum_{i=1}^{count_{AV}}w_2(i)}\text{ where}$$

$F_1=b_p(a_{A-st},a_{FP})$;

w_p is free parameter.

$F_2(i)=b_{AV-resp}(i)$ (This is the value of green error bar center);

$w_2(i)=(1-w_p)\cdot f_{wdt}(a_{A-st},a_{AV-st}(i),w_{wdt})$;

$count_{AV}=3$.

5.5.2 Performance

For spatially congruent AV stimuli the model looks like Figure 18. It was fitted by 'nlinfit' Matlab nonlinear regression fitting function.

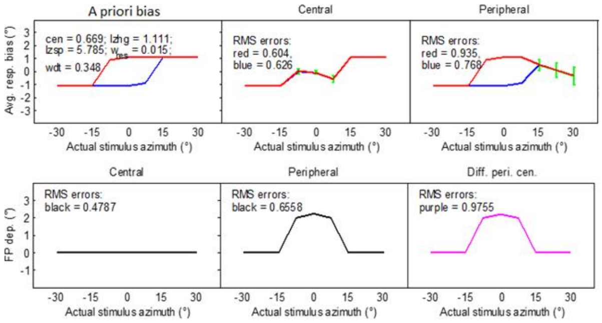

Figure 18: Modelling results

Resulting coefficients for fitting on no-shift data:

$c=0.669$,

$hg=1.111$,

$sp=5.785$,

$w_p=0.015$,

$w_{wdt}=0.348$;

Experimental data in Figure 16 are well-explained by current model. You can see that for central adaptation the red and blue lines are almost equal and this is also the case for prediction according to current model (Figure 18). You can also see that difference in red vs. Blue line for peripheral adaptation is present in central 3 azimuths for both experimental data and prediction. Magenta line is also similar for experimental data and prediction according to model.

But the prediction of magenta line according to previous model (Loksa and Kopco, 2016) would be zero constant for no-shift condition.

The model is less successful when we use fitting of also experimental data other than no-shift ones (Figure 14 and Figure 15). We did it and here are the results (Figure 19 and Figure 20).

The problem is that 5th row of these graphs, which displays difference of rows 2 and 3 divided by two, gives no difference between red and blue line, and that is inconsistent with experimental data (Figure 14 and Figure 15). This result shows the model's limitation in that the FP-specific shift (i.e., the difference between the red and blue lines in Figure 19) is independent of the direction or magnitude of the visually-induced adaptation. Thus, the model cannot describe any adaptation that is eye-centered. This can be proven analytically, as shown in the following section.

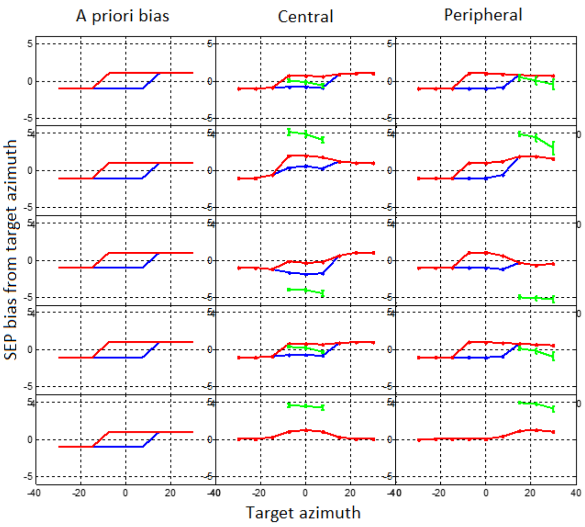

Figure 19: Prediction of no-shift, positive shift, negative shift, mean across shifts, mean across shifts oriented as positive ones, respectively for given rows. Prediction was done on experimental data from Figure 14 (Variables are in °).

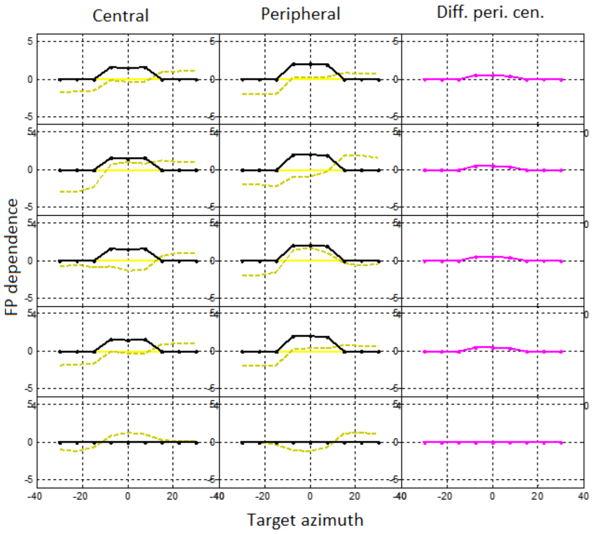

Figure 20 Continuation of previous figure.

Resulting coefficients for fitting on data on all 3 shift conditions:

$$c = 8.980,$$
$$hg = 0.989,$$
$$sp = 8.930,$$
$$w_p = 0.260,$$
$$w_{wdt} = 4.939;$$

Mean square error of model fitted on all 3 shift conditions:

$$\text{MSE} = 1.375;$$

5.5.3 Proof of the inappropriateness of the current model

This section shows the weak aspect of current model in the manner that proves its weakness algebraically.

Model formula (bias to A-only response):

$$b_{A-resp}(a_{A-stim}, a_{FP}, a_{AV-stim}, b_{AV-resp}) = \\ = \frac{F_1 \cdot w_p + \sum_{i=1}^{count_{AV}} F_2(i) \cdot w_2(i)}{w_p + \sum_{i=1}^{count_{AV}} w_2(i)}$$

Main substitution I. (frequent variables merged):

$$o = \{a_{A-st}, a_{AV-st}\}$$

Main substitution II. (FP-dependence, difference of bias between different FP-s):

$$d_{FP}(a_{FP1}, a_{FP2}, bias_{AV-resp}, o) = \\ = b_{A-resp}(o.a_{A-st}, a_{FP1}, o.a_{AV-st}, b_{AV-resp}) - \\ - b_{A-resp}(o.a_{A-st}, a_{FP2}, o.a_{AV-st}, b_{AV-resp})$$

Main hypothesis (FP-dependence independent of shift, thus also independent of $a_{AV-resp}$):

$$\forall(g, h, a_{FP1}, a_{FP2}, o): \\ :(d_{FP}(a_{FP1}, a_{FP2}, a_{AV-resp-g}, o) = \\ = d_{FP}(a_{FP1}, a_{FP2}, a_{AV-resp-h}, o))$$

5.5.3.1 Inference

Substitution 1 (denominator):

$$B_d(o) = w_p + \sum_{i=1}^{count_{AV}} w_2(i, o.a_{A-st}, o.a_{AV-st}(i))$$

Substitution 2 (first member of numerator (a priori)):

$$B_p(a_{FP}, o) = F_1(o.a_{A-st}, a_{FP}) \cdot w_p$$

Substitution 3 (second member of numerator (AV effect)):

$$B_{AV}(b_{AV-resp}, o) = \sum_{i=1}^{count_{AV}} F_2(i, b_{AV-resp}(i)) \cdot w_2(i, ...)$$

Inference 1: (Equivalent notation of model formula. To confirm, substitute members of current formula and compare)

$$b_{A-resp}(a_{A-stim}, a_{FP}, a_{AV-stim}, b_{AV-resp}) =$$
$$= \frac{B_p(a_{FP}, o) + B_{AV}(b_{AV-resp}, o)}{B_d(o)}$$

Inferences 2 and 3 (Derived from main substitution II. and result of inference 1.) We see that d_{FP} is independent from $b_{AV-resp}$ variable in following two formulas (so independent of shift):

$$d_{FP}(a_{FP1}, a_{FP2}, b_{AV-resp}, o) =$$
$$= \frac{B_p(a_{FP1}, o) + B_{AV}(b_{AV-resp}, o)}{B_d(o)} -$$
$$- \frac{B_p(a_{FP2}, o) + B_{AV}(b_{AV-resp}, o)}{B_d(o)}$$

Thus:

$$d_{FP}(a_{FP1}, a_{FP2}, b_{AV-resp}, o) =$$
$$= \frac{B_p(a_{FP1}, o) - B_p(a_{FP2}, o)}{B_d(o)};$$

Main hypothesis contradiction:

$$\exists(g, h, a_{FP1}, a_{FP2}, a_{A-stim}):$$
$$:[d_{FP}(a_{FP1}, a_{FP2}, b_{AV-resp-g}, o) \neq$$
$$\neq d_{FP}(a_{FP1}, a_{FP2}, b_{AV-resp-h}, o)]$$

Contradiction inference: (by substituting main hypothesis contradiction by result of inference 3)

$$\exists(g, h, a_{FP1}, a_{FP2}, azi_{stim}):$$
$$(\frac{B_p(a_{FP1}, o) - B_p(a_{FP2}, o)}{B_d(o)} \neq$$
$$\neq \frac{B_p(a_{FP1}, o) - B_p(a_{FP2}, o)}{B_d(o)})$$

Contradiction of main hypothesis is disproved, thus the main hypothesis, that so called FP-dependence is independent of shift direction ($b_{AV-resp}$ depends on the shift direction), is proved. This fact can be visually seen on Figure 20: 2nd, 3rd and 5th row.

5.6 Conclusion

We have described previous studies examining the reference frame of the ventriloquism aftereffect and its main results, which contain some ambiguity. We examined a part of the experimental data from that study, and we described a new adaptive phenomenon. We made the attempt to model these data and we have proven that the proposed model of a priori bias affected by AV responses seems appropriate to explain newly observed phenomenon when looking to the no-shift conditions, but inappropriate for the explanation of the difference of reference frames for shifted conditions (Figure 13 and Figure 15).

One of the alternatives for current modelling is use different weights for TrFP vs. NtrFP. Other one is to make NtrFP biases depending on TrFP biases instead directly of AV biases. Alternatively, completely other factors might play role. Additional modeling is currently required to examine these alternatives.

5.7 Acknowledgement

This contribution was supported by Slovak Research and Development Agency under grant project APVV-0452-12 Spatial Attention and Listening in Complex Environments and by the TECHNICOM project, ITMS: 26220220182, of the EU RDP.

6 A model of the reference frame of the ventriloquism aftereffect

The current chapter contains the content of the article: Loksa and Kopco (2021). All work described here was performed by the thesis author under the supervision of the thesis advisor.

6.1 ABSTRACT

Background: Ventriloquism aftereffect (VAE), observed as a shift in the perceived locations of sounds after audiovisual stimulation, requires reference frame (RF) alignment since hearing and vision encode space in different RFs (head-centered, HC, vs. eye-centered, EC). Experimental studies examining the RF of VAE found inconsistent results: a mixture of HC and EC RFs was observed for VAE induced in the central region, while a predominantly HC RF was observed in the periphery. Here, a computational model examines these inconsistencies, as well as a newly observed EC adaptation induced by AV-aligned audiovisual stimuli.

Methods: The model has two versions, each containing two additively combined components: a saccade-related component characterizing the adaptation in auditory-saccade responses, and auditory space representation adapted by ventriloquism signals either in the HC RF (HC version) or in a combination of HC and EC RFs (HEC version).

Results: The HEC model performed better than the HC model in the main simulation considering all the data, while the HC model was more appropriate when only the AV-aligned adaptation data were simulated.

Conclusion: Visual signals in a uniform mixed HC+EC RF are likely used to calibrate the auditory spatial representation, even after the EC-referenced auditory-saccade adaptation is accounted for.

6.2 Introduction

Auditory spatial perception is highly adaptive and visual signals often guide this adaptation. In the "ventriloquism aftereffect" (VAE), the perceived location of sounds presented alone is shifted after repeated presentations of spatially mismatched visual and auditory stimuli (Recanzone, 1998b; Woods and Recanzone, 2004; Bertelson *et al.*, 2006). Complex transformations of spatial representations in the brain are necessary for the visual calibration of auditory space to function correctly, as visual and auditory spatial representations differ in many important ways. Here, we propose a computational model and perform a behavioral data analysis to examine the visually guided adaptation of auditory spatial representation in VAE and the related transformations of the reference frames (RFs) of auditory and visual spatial encoding.

Several previous models were developed to describe the ventriloquism aftereffect in humans and birds. The bird models examined VAE in the barn owls (Haessly *et al.*, 1995; Oess *et al.*, 2020) which cannot move their eyes and therefore do not need to re-align the auditory and visual RFs. The human models mainly focused on spatial and temporal aspects of the ventriloquism aftereffect (Shinn-Cunningham *et al.*, 2005; Bosen *et al.*, 2018; Watson *et al.*, 2019), not considering the differing RFs. There are models of the audio-visual reference frame alignment, but those only consider audio-visual integration (Razavi *et al.*, 2007) and multi-sensory integration (Pouget *et al.*, 2002) when in the auditory and stimuli are presented simultaneously, like in the ventriloquism effect, not the adaptation and transformations underlying VAE.

Here, we primarily examine the reference frame (RF) in which VAE occurs. While visual space is initially encoded relative to the direction of the eye gaze, the cues for auditory space are computed relative to the orientation of the head (Groh and Sparks, 1992). A means of aligning these RFs is necessary by the stage at which the visual signals guide auditory spatial adaptation. Our previous studies suggest that a mixture of eye-centered and head-centered RFs is associated with recalibration in the central region of the audiovisual field (Kopco *et al.*, 2009) while the head-centered RF dominates for VAE locally induced in a single hemifield in the visual periphery (Kopco *et al.*, 2019b). These results imply that the RF used in VAE is location dependent, possibly due to non-homogeneity in the auditory spatial representation. Specifically, recent evidence suggests that, in mammals, auditory space encoding is based on two or more spatial channels roughly aligned with the left and right hemifields of the horizontal plane (Grothe *et al.*, 2010; Groh, 2014). The current modeling explores an alternative hypothesis about the location-dependence of the RF of VAE. It assumes that the RF transformations are the same across the audio-visual field, and that the observed location-dependence is due to other adaptive processes, e.g., related to auditory saccade adaptation, as saccades were used to measure behavioral responses in the Kopco et al. (Kopco *et al.*, 2009; Kopco *et al.*, 2019b) studies. The main modeling goal is then to determine whether such a uniform, location-independent spatial adaptation is only driven by head-orientation referenced visual signals, or whether signals in eye-centered RF also contribute.

The second question explored here is how to separate the effect of auditory saccade adaptation from the ventriloquism-induced auditory space adaptation. Previous studies show that auditory saccades can overestimate or underestimate the actual sound locations (Yao and Peck, 1997) and that the amount of visually induced adaptation does not depend on whether the resulting saccades are hypometric or hypermetric (Pages and Groh, 2013). Here, in the Appendix, we analyze the data from Kopco et al. (Kopco *et al.*, 2009; Kopco *et al.*, 2019b) to determine whether the ventriloquism effect and aftereffect show asymmetries depending on the resulting adaptation type (hypometric vs. hypermetric), as well as on the saccade amplitude magnitude. Based on this analysis, the current model assumes that the magnitude of the ventriloquism aftereffect is proportional to the magnitude of the ventriloquism effect, independent of whether these shifts result in hypometric or hypermetric saccades, and independent on the saccade magnitude.

Finally, Kopco et al. (Kopco *et al.*, 2019b) observed a new adaptive phenomenon induced by aligned audiovisual stimuli presented in the periphery, exhibited as a shift in responses to sounds presented alone in the central region. The shift magnitude depended on the gaze direction and, thus, was at least partly in the eye-centered RF. However, no such shift was observed when aligned audiovisual stimuli were presented in the central region (Kopco *et al.*, 2009). The current model proposes a mechanism of a priori biases in the saccade responses, possibly due to auditory saccade adaptation, that can describe this phenomenon.

In the paper, we first summarize the Kopco et al. (Kopco *et al.*, 2009; Kopco *et al.*, 2019b) data modeled here, and, in the Appendix, provide a new analysis of these data to examine 1) how VAE magnitude depends on whether it results in hypometric vs. hypermetric saccades, and 2) how the VAE magnitude relates to the magnitude of the ventriloquism effect. Then, the model is introduced and two versions of it are examined in 4 simulations, each focusing on different aspects of the data and model components. The main result of the simulations is that a common location-independent mechanism can describe the data best when visual signals adapt the auditory spatial map in both head-centered and eye-centered reference frames, consistent with the idea that the reference frame of ventriloquism aftereffect is mixed.

6.3 Experimental data

This section summarizes the experimental methods and results from Kopco et al. (Kopco *et al.*, 2009; Kopco *et al.*, 2019b). Additionally, Appendix presents results of a new analysis of the data aimed at examining the dependence of the results on the properties of auditory saccades used by subjects for responding.

In the experiments, ventriloquism was induced by audio-visual training trials either in the central or peripheral subregion of the horizontal audio-visual field while the eyes fixated one location (red '+' symbol; upper and middle panels of Figure 21(A)). The aftereffect was evaluated on interleaved auditory-only probe trials using a wide range of target locations while the eyes fixated one of two locations (lower panel of Figure 21(A)). The listener's task in both audio-visual and auditory-only trials was to perform a saccade to the perceived location of the auditory stimulus/component from the FP. It was expected that the AV stimuli with displaced visual component would induce a local ventriloquism aftereffect when measured with the eyes fixating the training FP (red dash-dotted lines in Figure 21(B) illustrate this prediction for the peripheral-training experiment). Confirming this expectation, the red solid and dashed lines in Figure 21(B) show that maximum ventriloquism was induced in the peripheral and central training subregion, respectively. The critical manipulation of these experiments was that a subset of probe trials was performed with eyes fixating a new, non-training fixation point (blue '+' symbol), located 23.5° to the left of the training fixation. As illustrated by the blue dash-dotted line in Figure 21(B), if the RF of VAE is purely head-centered, then moving the eyes to a new location is expected to have no effect, resulting in the same pattern of ventriloquism for the non-training and training FPs. On the other hand, if the RF is purely eye-centered, the observed pattern of induced shifts is expected to move with the eyes when the eyes are moved to a new location, as illustrated by the cyan dash-dotted line. The experimental data showed that, in the central experiment, moving the fixation resulted in a smaller ventriloquism aftereffect with the peak moving in the direction of eye gaze (blue dashed line), while in the peripheral experiment no effect of eye gaze position was observed (blue solid line). To better visualize these results, the lower panels of Figure 21(B) shows predictions and data expressed as difference between responses from training vs. non-training FPs from the respective upper panels. The head-centered RF always predicts that the effect would be identical for the two FPs. Thus, all head-centered predictions (brown lines) are always at zero. The yellow dash-dotted line shows a hypothetical prediction for eye-centered RF, obtained by subtracting the cyan from the red dash-dotted line. Similarly, the solid and dashed yellow lines show, respectively, for the peripheral and central data, the eye-centered RF predictions obtained by subtracting from the red lines the same red lines shifted 23.5° to the left. Finally, the black solid and dotted lines show the actual differences between the respective red and blue data from the upper panels. For the central data, the black dashed line falls approximately in the middle between the head-centered and eye-centered predictions, showing a mixed nature of the RF of VAE induced in this region. On the other hand, the black solid line is always near zero, confirming that the RF of VAE induced in the periphery is predominantly head-centered. The current model aims to describe these differences by considering a uniform representation and adaptation process that guided by signals in both eye-centered and head-centered reference frames.

The results described in Figure 21(B) are based on ventriloquism aftereffect induced by visual stimuli displaced to the left or to the right of the corresponding auditory stimuli. Figure 21(C) shows the baseline data obtained in runs with auditory and visual stimuli aligned. In the central-training experiment, the responses from the two FPs were similar, unbiased at the central locations and with a slight expansive bias in the periphery (both red and blue dotted lines are near zero in the center, negative in the left-hand portion and positive in the right-hand portion of the graph). On the other hand, in the peripheral-training experiment the responses in the central region differed between the two fixations, where the non-training FP responses fell well below the training-FP responses (compare the red and blue solid lines). Thus, the peripheral AV-aligned stimuli induced a fixation-dependent adaptation in the auditory-only responses in the central region. The black dashed and solid lines in Figure 21(C), showing the difference between the corresponding training and non-training FP data, highlight the FP-dependence of the peripheral experiment in contrast to the FP-independence in the central experiment. The current model assumes that these adaptive effects can be explained by a combination of biases in visual saccades to auditory stimuli and a visually guided adaptation in the spatial auditory representation.

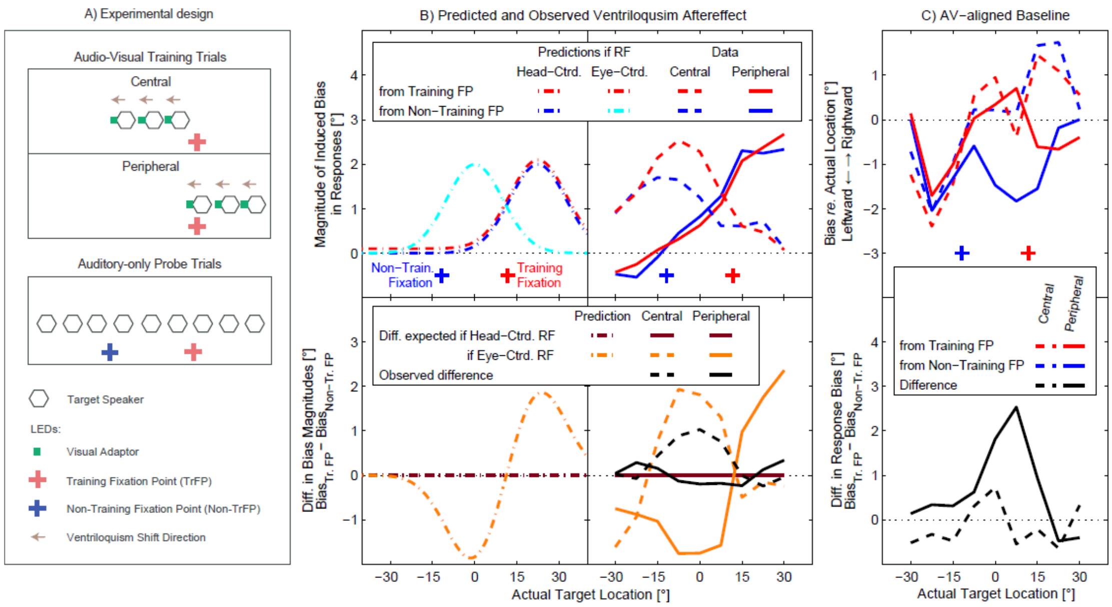

Figure 21: Experimental design and predicted and observed ventriloquism aftereffect from Kopco et al. (Kopco *et al.*, 2009; Kopco *et al.*, 2019b). A) Setup: nine loudspeakers were evenly distributed at azimuths from -30° to 30°. Two fixation points were used, located 10° below the loudspeakers at +-11.25°. On training trials, audiovisual stimuli were presented either from the central region (Kopco *et al.*, 2009) or peripheral region (Kopco *et al.*, 2019b), while the subject fixated one FP. The audiovisual stimuli consisted of a sound paired with an LED offset by -5°, 0°, or -5° (offset direction fixed within a session). On interleaved probe trials, the sound was presented from any of the loudspeakers while the eyes fixated either one of the FPs. B) Predicted (left-hand panels) and observed (right-hand panels) reference frames of the ventriloquism aftereffect. Lines represent model predictions or across-subject means of the aftereffect magnitudes for the probe trials from the AV-misaligned runs. C) Across-subject mean aftereffect magnitudes for the probe trials from the AV-aligned runs. Note: Error bars have been omitted for clarity. They are presented in the simulation figures in which data are compared to models.

6.4 Model Description

6.4.1 Overview

Figure 22A shows the outline of the model. The model predicts the azimuthal bias in the saccade response to an auditory-only probe (the “Response” block in panel A) as a function of the probe azimuth, with additional parameters of the fixation location on a given trial (“Probe stimulus and fixation” block) and the audio-visual training locations and the measured audio-visual response biases in a given experimental training session (“Ventriloquism” block). Thus, the model does not require information about the direction of audio-visual stimulus displacement during training (whether the visual stimuli were shifted to the left, right, or aligned with the auditory stimuli). Instead, it only uses the information about where the training occurred and what the resulting ventriloquism effect was. Here, the model assumes that there is a direct relation between the observed ventriloquism effect and aftereffect, as shown in the Appendix. The ventriloquism aftereffect prediction is then modeled as an additive combination of two components, a saccade-related bias in eye-centered reference frame and a saccade-independent visually guided adaptation of auditory space representation (square blocks in panel A). The saccade-related bias is present *a priori* and it is not directly adapted by ventriloquism, while the auditory spatial representation is locally adapted by the ventriloquism signals in different reference frames and its size also depends on the saccade-related bias.

Two versions of the model are evaluated, differing only by the assumed form of adaptation of the auditory space representation. First, in the HC model, the visual signals adapt the auditory spatial representation exclusively in the head-centered reference frame (the “HC” arrow in panel A), so the signals are assumed to be transformed to HC before inducing adaptation. In the HEC model, the visual signals adapt the auditory spatial representation in both head-centered and eye-centered RFs (“HC” and “EC” arrows) such that the relative contribution of the HC and EC RFs can be arbitrary. I.e., the HEC model reduces to the HC model if the weight of the EC path is set to zero, or it can produce predictions using only EC RF if the HC weight is set to zero.

In summary, both models assume that the spatial representations and adaptations are uniform, predicting the same results independent of whether the training occurs in the center or in the periphery. The main difference between the two models is that the HC model assumes that the auditory space adaptation occurs purely in head-center coordinates, while it is the gaze-direction-referenced properties of the auditory saccades that cause any eye-centered effects observed in the data. On the other hand, the HEC model assumes that, even after accounting for the saccade-related effects, the auditory spatial representation receives the adaptive visual signals in both reference frames, causing adaptation that always depends on the position of the stimuli relative to the eye gaze direction. Importantly, the model assumes that if the ventriloquism aftereffect is not induced and measured by auditory saccades, as used in the Kopco et al. (Kopco *et al.*, 2009; Kopco *et al.*, 2019b). studies, the saccade-related bias would not affect the performance.

6.4.2 Detailed Specification

The following model specification applies to the more general HEC model version, with the differences applying to the HC model described as needed. Panels B-D of Figure 22 provide visualizations of the behavior of different parts of the model.

Equation 1 describes the predicted bias in responses $\hat{r}$ to a given auditory stimulus location s as a weighted sum of a saccade-related bias r_E and a ventriloquism-related adaptation in auditory spatial representation r_V

$$\hat{r}(s) = r_E(s) + w \cdot r_V(s), \tag{1}$$

where $w \in [0, \infty]$ is a free parameter specifying the relative weight of the ventriloquism adaptation. In addition to the stimulus location s, the prediction (illustrated in Fig. 3D) also depends on the fixation point on a given trial f, on the training region specified by the training AV stimulus locations s_{AV}, and on the observed biases in AV stimulus responses at these locations r_{AV} (all variables in the units of degrees).

The saccade-related bias at a specific location x for eyes fixating the location f is modeled as a sigmoidal function

$$r_E(x) = \frac{2 \cdot h}{1 + \mathrm{e}^{-k(x+cf)}} - 1, \tag{2}$$

where $h, k,$ and c are free parameters characterizing the sigmoid. The saccade-related bias (Figure 22B) is broad and referenced to the FP (i.e., it uses EC RF), exhibiting a combination of underestimations and overestimations commonly observed in studies of auditory saccades (Yao and Peck, 1997; Razavi *et al.*, 2007; Gabriel *et al.*, 2010). However, the specific shape of the functions used here was chosen to best fit the peripheral and central no-shift data shown in Fig. 1C. Specifically, the predictions roughly follow the values observed at each location in Fig. 1C when no audiovisual training is used at a given location (the central-experiment data for the right-most location triplet, the peripheral-experiment data for the central triplet, and data from both experiment for the left-most triplet). Thus, it is assumed that this saccade-related bias is present *a priori*, independent of the induced ventriloquism. Also, it is assumed that the bias only depends on the probe location re. FP location, which, for the current data means that the bias graphs for training and non-training FPs are symmetrical about the origin with respect to each other (blue and red lines in Fig. 2B).

The ventriloquism-driven auditory space adaptation causes bias defined at location x, for eyes fixating the location f, and for ventriloquism induced at training locations s_{AV} and resulting in AV response biases r_{AV}, as a weighted sum:

$$r_V(x) = \sum_{i=1}^{N} w_{v,i}(x) \cdot \left[r_{AV,i} - r_E\left(s_{AV,i}\right)\right], \tag{3}$$

where N is the number of training locations ($N = 3$ for the current study), i is an index through these locations, $s_{AV,i}$ is the i-th training location azimuth, and $r_{AV,i}$ is the AV response bias observed at the i-th training location. The differences $r_{AV,i} - r_s\left(s_{AV,i}\right)$ represent the disparity between the AV response biases (green diamonds in Figure 22B) and the saccade-related bias (red/blue lines in Figure 22B) at the training locations. The disparities are shown in Figure 22C by the red and blue full diamonds. $w_{v,i}(x)$ is the strength with which the disparity at the i-th training location adapts the spatial representation at the location x. In the HEC model, this value is a weighted sum of the adaptation strengths in head-centered and eye-centered reference frames, defined as:

$$w_{v,i}(x) = (1 - w_E) \cdot w_{vH,i}(x) + w_E \cdot w_{vE,i}(x), \tag{4}$$

where $w_E \in (0, 1)$ is a parameter determining the relative weight of the EC reference frame vs. the HC RF (in the HC model, $w_E = 0$). Finally, $w_{vH,i}$ and $w_{vE,i}$ use normalized Gaussian functions centered at training locations as a measure of influence of the i-th training location on the target location x, in the two reference frames:

$$w_{vH,i}(x) = G\left(x, s_{AV,i}, \sigma_H\right), \tag{5}$$

$$w_{vE,i}(x) = G\left(x, s_{AV,i} + f - 11.25°, \sigma_E\right), \tag{6}$$

$$G(x, \mu, \sigma) = \frac{\mathcal{N}(x,\mu,\sigma)}{\sum_{i=1}^{N} \mathcal{N}(7.5 \cdot (i-2), 0, \sigma)}, \tag{7}$$

In Eqs. 5 and 6, the parameters σ_H and σ_E represent the width of the influence of the ventriloquism shift at individual training locations, separately for the two reference frames. $w_{vH,i}$ (Eq. 5) is always centered on the i-th training location in the HC RF, whereas $w_{vE,i}$ (Eq. 6) is centered on the i-th training location in the EC RF (for the training FP, the two RFs are aligned). Finally, the Gaussian functions are normalized (Eq. 7) such that the maximum $w_{vH,i}$ or $w_{vE,i}$ after summing across the three training locations is 1 (the normalization locations $7.5 \cdot (i-2)$ are specific for the current training and they need to be modified for other data with different training locations).

Figure 22C shows the operation of the ventriloquism adaptation. As mentioned above, the red and blue filled diamonds are the disparities at the individual training locations driving the adaptation in HC RF. The blue open diamonds are identical to the blue filled diamonds except that they are shifted to the left by the difference between the two FPs to illustrate how the eye gaze shift affects where the adaptation is expected to occur in the EC RF. The red and blue lines are then the resulting biases r_V for the two fixation locations, each corresponding to the sum of Gaussians centered at different training locations in the two RFs (and with widths defined by the σ's). Parameter w_E determines the relative weights of the peaks in the blue line corresponding to the open diamonds vs. those corresponding to the filled diamonds. In summary, the blue and red lines show how visually guided adaptation is local and RF-dependent, decreasing with distance from location at which AV stimuli were present in HC and EC RFs. It also shows that since adaptation causes shifts from the saccade-bias response locations towards AV response locations, if AV responses fall on saccade bias locations, no visually guided adaptation is predicted to occur.

Finally, Figure 22D shows that the model prediction is a sum of the saccade bias (from Figure 22B) and ventriloquism bias (Figure 22C) weighted by the parameter w (note that no scaling parameter is needed for the saccade bias as parameter h already can make this bias arbitrarily large).

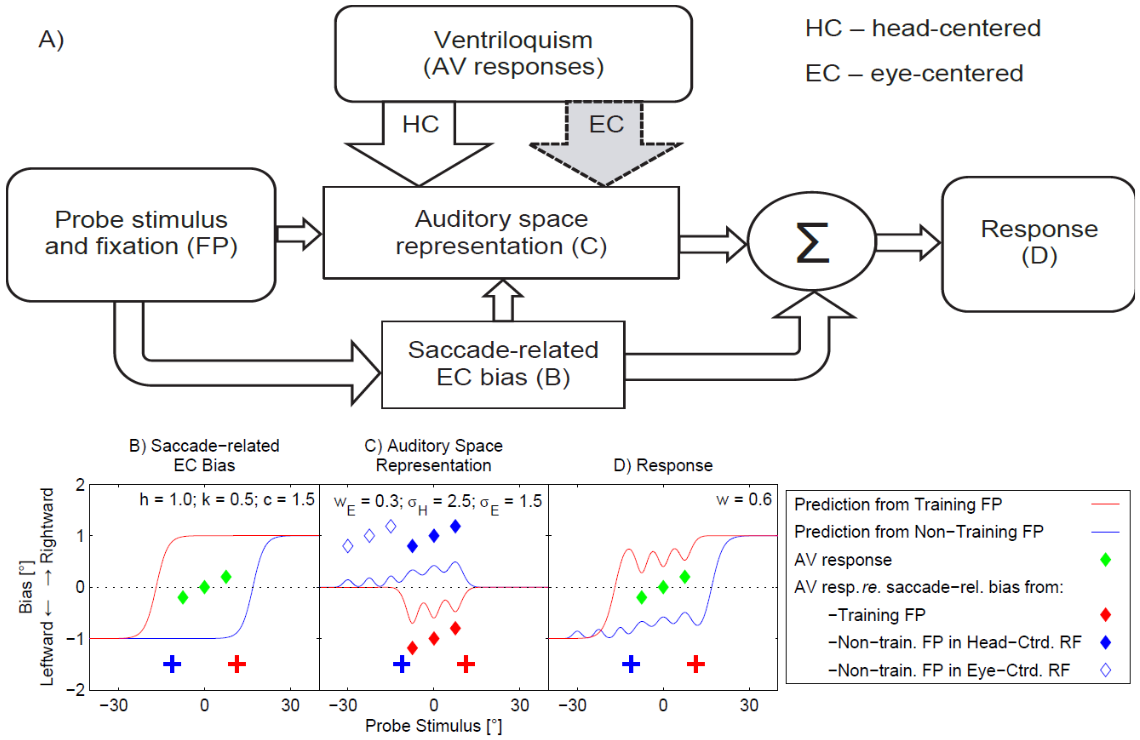

Figure 22: Structure of the HC/HEC model and illustration of its operation. A) Block diagram of the model. The model predicts the response bias as a function of the probe stimulus location, with additional input parameters of the fixation position, training locations, and the observed ventriloquism effect at the training locations (rounded blocks). Two mechanisms determine the response (square blocks). First, saccade-related bias is always present and it is not influenced by the ventriloquism signals. Second, auditory space representation which is adapted by ventriloquism only in HC reference frame (HC model; "HC" arrow) or in a combination of HC and EC RFs (HEC model; "HC" and "EC" arrow). Labels B, C, D within blocks refer to respective panels below that illustrate the function of the blocks by showing the outputs of the model components in an illustrative simulation (for training in the central region for which the observed AV responses are nearly unbiased). B) Saccade-related bias predictions for the two fixation points (red and blue lines). The green diamonds show the nearly zero ventriloquism effect assumed for the predictions shown in panel C. C) Adaptation of auditory space representation resulting from the saccade-related bias and AV response bias as shown in panel B. Diamonds represent the disparity between AV response bias and saccade-related bias for the training FP (red), and non-training FP in HC RF (blue filled) and in EC RF (blue open). Lines represent predictions of auditory space adaptation induced by these disparities. D) Response bias predicted by the model as a weighted combination of biases shown in panels B and C. Values of model parameters used for the predictions of respective model components are shown along the upper frame in each panel.

6.5 Methods

6.5.1 Stimuli

The data from studies of Kopco et al. (Kopco *et al.*, 2009; Kopco *et al.*, 2019b), simulated here, induced ventriloquism by presenting training stimuli with visual component either shifted to the left, to the right, or aligned with the auditory component, while the eyes fixated one location (Fig. 1A; upper and middle panels). The aftereffect was always measured by presenting auditory-only stimuli while eyes fixated one of the two FPs (Figure 21A; lower panel). Thus, nominally, there were 6 conditions (3 shift directions by 2 training regions), corresponding to AV locations and responses shown by triplets of open symbols in Figure 27A. For these conditions, predictions could be compared to data for 9 locations at 2 FPs. However, the main experimental results simulated here were observed when differences between FPs were considered on aftereffect magnitude data, obtained by subtracting positive-shift data from negative-shift data and halving the result (Figure 21B; lower right panel; note that the latter difference is equivalent to averaging the magnitudes of "positive shift – no shift" and "negative shift – no shift"). These "double differential" ("positive – negative" difference of "training FP – non-training FP" difference) data were the most stable as they eliminated a lot of between-subject variability related to individual biases in responses (as will be illustrated later). Therefore, to focus the model on these important differences, the data were also transformed into the difference representation in two steps.

First, the data for the two training FPs were orthogonally transformed such that instead of using training and non-training FP, a sum and a difference across the two FPs was used. I.e., instead of having for each condition 18 data points corresponding to 9 locations at 2 FPs, we used 18 data points consisting of 9 locations summed across the two FPs and 9 locations for difference across the 2 FPs.

Second, the positive-shift and negative-shift condition data were transformed in a similar way, such that instead of positive and negative shift we used the aftereffect magnitude (i.e., a halved difference between the two shifts) and average across the two shifts. The no-shift data were left unmodified.

The complete data set therefore consisted of 108 data points [9 (locations) x 2 (transformed FPs) x 3 (transformed shifts) x 2 (training regions)]. Across-subject mean and standard deviation data were used in the simulations.

6.5.2 Simulations

Four simulations were performed in this study, each assessing both the HC and HEC models on a different subset of the Kopco et al. (Kopco *et al.*, 2009; Kopco *et al.*, 2019b) data. The first two simulations, No-Shift and All Data simulations, tested two main hypotheses about the current data and reference frame. Two supplementary simulations, Central Data and Peripheral Data simulations, were performed confirm that the model behavior matches the conclusions of the Kopco et al. (Kopco *et al.*, 2009; Kopco *et al.*, 2019b) studies when considered separately.

No Shift simulation assessed the models on the AV-aligned baseline no-shift data from both experiments (Figure 21C), examining the interaction between the saccade-related bias and visual signals when no ventriloquism is induced.

All Data simulation is the main simulation of this study. In this simulation the models were fitted on the complete dataset from both experiments (Figure 21B and C) to examine whether a uniform representation of the reference frame of ventriloquism aftereffect is mixed or purely head-centered.

Central Data simulation fitted only the central-training data from the positive-shift and negative-shift conditions (dashed lines in Figure 21B) while predictions were generated for all the data. The main goal was to examine the reference frame in which the ventriloquism aftereffect is induced in the central region.

Peripheral Data simulation fitted only the peripheral-training data from the positive-shift and negative-shift conditions (solid lines in Figure 21B) while predictions were generated for all the data. The main goal was to examine the reference frame in which the ventriloquism aftereffect is induced in the audiovisual periphery.

6.5.3 Model Fitting and Evaluation

Each simulation was performed by fitting the two models to the corresponding subset of the transformed data using a two-step procedure. First, a systematic search through the parameter space was performed, using all combinations of 10 values for each parameter, listed in Table 2 (HEC model used all 7 parameters, while HC model only used 5 of them). The limits of the range were chosen by piloting to cover the expected range of behaviors of the model. Note that quadratic spacing was chosen for parameters k and c as the behavior of the sigmoidal function varies non-uniformly with the parameter values (k was sampled more densely at the lower end of the range, c at the higher end). Then we selected the best 100 parameter combinations in terms of weighted MSE, in which each data point was weighted by the inverse of the across-subject standard deviation in that data point. These parameter combinations were then used as starting positions for non-linear iterative least-squares fitting procedure (Matlab function lsqnonlin) which, again, minimized the weighted MSE. The parameter values obtained by the best of these fits were chosen as the optimal values.

Table 2: The range and increments of values of free parameters used in systematic search through the parameter space during model simulations. Ten values of each parameter were considered with either linear or quadratic spacing. Note that parameters w_E and σ_E are not used in the HC model, while all parameters are used in the HEC model.

Parameter	Range		Increments
	min	max	
h, w	0	2	linear
k	0.01	20	quadratic
c	0	1.5	quadratic
w_E	0	1	linear
σ_H, σ_E	1	20	linear

To compare the models' performance while accounting for the number of parameters used by each model, we computed the Akaike information criterion AICc (Burnham and Anderson, 2004; Taboga, 2017) for each optimal fit, defined as:

$$\text{AICc} = -2\log(L) + 2K + 2K\frac{K+1}{n-K-1}, \tag{8}$$

$$\log(L) = -\frac{n}{2}\left(\log(2\pi)+\log\frac{\text{SSE}(X)}{n} + 1\right) \tag{9}$$

where n is the number of experimental data points, K is the number of fitted parameters, and $\text{SSE}(X)$ is the sum of squares of errors across the data points (i.e., differences between predictions and across-subject mean data x_i) weighted for each data point by the inverse of its across-subject standard deviation $\frac{1}{SD(x_i)}$. In general, the model with the lower AICc is considered to be a better fit for the data. Then, to determine whether the data provide substantial support for one model over the other one, we computed ΔAIC as the difference in AICc values of the model with the higher AICc vs. the one with the lower AICc. And, we use the following rule to determine whether the model with the lower AICc is substantially better than the other model (Burnham and Anderson, 2004): “Models having $\Delta\text{AIC} \leq 2$ have substantial support (evidence), those in which $4 \leq \Delta\text{AIC} \leq 7$ have considerably less support, and models having $\Delta\text{AIC} > 10$ have essentially no support.”. Thus, only if ΔAIC is substantially larger than 2, the result is interpreted as evidence in favor of the model with lower AICc.

6.6 Simulation Hypotheses and Results

The results of the 4 simulations performed in this study are summarized in Table 3, which shows for each simulation and model the fitted model parameter values and the model’s performance measured using the AICc criterium.

Table 3: Values of fitted model parameters and evaluation of model performance for each simulation. AICc states the criterion for a given simulation, ΔAIC is the increase in AICc for a given simulation *re.* the simulation on a given data with the minimum AICc. The underscored model names indicate the model for which there is a substantial evidence of being a better fit for the data (rounded up value of ΔAIC smaller than 2).

Simulation	Model	Fitted parameter values							Performance	
		h	k	c	w	w_E	σ_H	σ_E	AICc	ΔAIC
No Shift (Figure 23)	HC	1.03	0.31	1.14	1.01	-	12.06	-	130.90	2.36
	HEC	1.13	0.17	0.95	1.24	0.36	12.84	2.98	128.54	-
All Data (Figure 24)	HC	0.79	0.82	1.15	0.49	-	14.21	-	444.75	7.86
	HEC	0.77	0.76	1.13	0.53	0.15	13.35	4.83	436.89	-
Central (Figure 25)	HC	1.01	5.64	0.67	0.40	-	18.79	-	176.16	5.92
	HEC	0.96	5.60	0.67	0.48	0.30	18.14	5.01	170.24	-
Peripheral (Figure 26)	HC	0.83	3.40	1.33	0.55	-	12.43	-	136.33	-
	HEC	0.82	5.33	1.33	0.56	0.04	12.12	4.91	141.89	5.56

6.6.1 No-shift simulation

This simulation focused on the AV aligned data, examining the hypothesis that *the saccade-related bias combined with auditory space adaptation in HC RF causes the* *training-region-dependent differences in the AV-aligned baseline data* (Figure 21C). I.e., it was predicted that EC visual signals adapting the auditory space representation do not need to be considered to explain the different adaptation effects observed in central vs. peripheral AV-aligned training. This hypothesis would be confirmed if the two models, HC and HEC, captured the behavioral data equally well.

Figure 23 presents the results of the simulation of the AV-aligned baseline no-shift condition from both experiments. Panel A shows the biases of the two model components (rows) for each of the two models (colors) with the fitted parameters as listed in Table 3, separately for the two fixation points (columns). The same fitted model parameters apply to both the central and peripheral training experiments. For the saccade-related bias (upper row) that means that the plotted graphs apply to both data equally. However, for the auditory space adaptation component (lower row), the plotted graphs apply to central training, since they show the effect of training at the 3 central locations (-7.5°, 0°, +7.5°). The graphs need to be shifted to the right by 22.5° to see their effect for peripheral training data.

Panel B shows the data (circles with error bars corresponding to the standard error of the mean) and predictions of the two models (lines), separately for the two training points (upper and middle rows), as well as for the difference between the FPs (lower row). The columns represent the two training regions. Each prediction in the upper and middle rows is, roughly, a weighted sum of the corresponding components from panel A, while the predictions in the lower row of panel B show the differences of the predictions from the upper and middle rows.

Considering the model predictions of the mean data, both models captured all the significant trends in these data. Specifically, for the central training data, both models predicted the slight expansion of the space for the central training data identical for both FPs (upper and middle row of the left-hand column), as well as the FP-dependence of the peripheral training data at the central locations (upper and middle row of the right-hand column). Most importantly, both models captured very well the difference data, which are near zero for the central training experiment and have a positive deviation for the peripheral training (bottom row). This conclusion is confirmed by the AICc evaluation which showed no evidence that either of the models should be preferred ($\Delta \mathrm{AIC} = 2.4$).

The data in panel B are replotted from Figure 21C, now also including the error bars. These error bars show that there was a lot of across-subject variability when the individual

FPs were considered (upper and middle row), while a large portion of that variability was eliminated when the differences in biases across the FPs were computed (lower row). This illustrates why the models were fitted on the transformed data, as those were much more consistent across subjects, and, with the transformation, the fitting weighed the difference data (lower row) more as they were much more reliable. Note that the second transformed data set, the average across FPs, is not shown, as it can be easily estimated from the individual FP data in the upper two rows of panel B.

Panel A illustrates how the models achieved the correct prediction. Both models predicted similar saccade-related bias, consisting of expansion at the peripheral target locations (+/-15°, +/-22.5°, and +/-30°) and bias towards the fixation location for the central 3 locations (upper row). This saccade-related bias was then modulated by the auditory space adaptation such that at the training locations the model predictions were shifted towards the AV responses, which were near zero for both the central and peripheral training (Figure 27A). The HC model predicts that this "corrective" ventriloquism shift only occurred in HC RF (brown lines in the lower row of panels), while the HEC model predicts a considerable contribution of the EC RF (magenta lines at locations -30° to -15° at the bottom right). However that contribution only had a small effect on the overall predictions, as shown by the small differences between the brown and magenta lines in panel B.

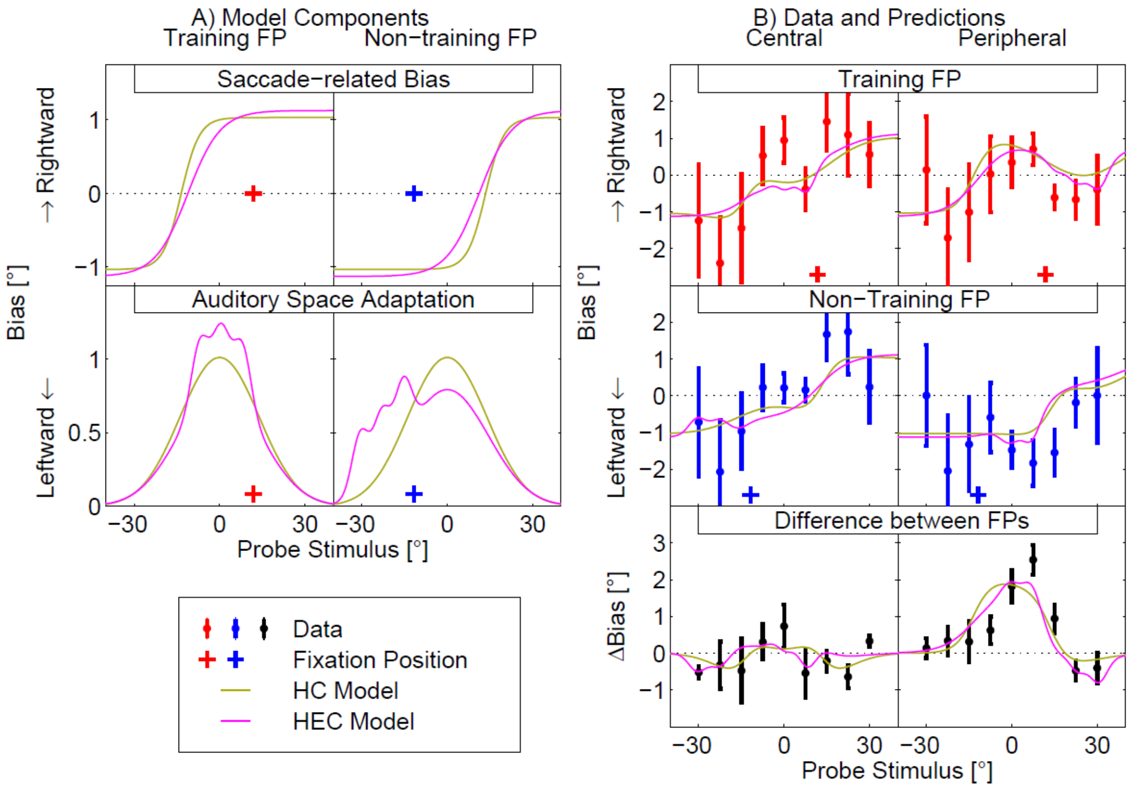

Figure 23: Model predictions and data for the No-Shift simulation. A) Visualization of the two model components, Saccade-Related Bias and Auditory Space Adaptation, for the HC and HEC models with the parameters fitted to the no-shift data (from Table 3). The Saccade-related Bias component (upper row) is independent of any visually guided ventriloquism adaptation. The Auditory Space Adaptation component (lower row) shows the strength with which the ventriloquism induced by the AV stimuli at 3 central locations shifts the responses from the Saccade-Related Bias locations to the AV response locations (Eq. 3). Note that for peripheral-training data, i.e., for the AV stimuli at the locations of 15°-30°, the lower-row graphs would be shifted by 22.5° to the right. B) Across-subject mean biases (±standard error of the mean) and model predictions for the two fixation locations (upper and middle row) and the difference between the two fixations (lower row).

6.6.2 All Data simulation

This was the main simulation of this study. The two models were fitted on the positive-shift and negative-shift data, in addition to the no-shift data from the previous simulation (Figure 21B and C). Also, the simulation was performed on the data from both experiments. Thus it assumed that the reference frame of ventriloquism aftereffect is uniform across the audiovisual field, as the models were optimized to fit both the central and peripheral training data simultaneously. The simulation further assumed that the saccade-related component of the model accounts for all the saccade-related effects (which are EC-referenced), an assumption supported by the results of the No Shift simulation. With these assumptions, the simulation examined the hypothesis that the *RF is mixed, using visual signals in both head-centered and eye-centered coordinates*. This hypothesis would be confirmed if the HEC model, using both HC and EC referenced visual signals, captured the behavioral data significantly better than the HC model, which only uses HC RF for the ventriloquism adaptation of the auditory space.

Figure 24 presents the results of this simulation. Panel A shows the biases of the two model components for the fitted parameter values from Table 2, in a format similar to panel A of Figure 23. Panel B shows the data (circles with error bars corresponding to the standard error of the mean) and predictions of the two models (lines). Panel B shows for this simulation only the difference of Training vs. Non-training FP data, equivalent to the black lines in Figure 21B and 1C. The upper row of panel B shows the no-shift data replotted from Figure 21C (also shown in the bottom row of Figure 23B), while in the lower row shows the difference between the positive-shift and negative-shift data, equivalent to a doubling of the aftereffect magnitude data from Figure 21B (black solid and dashed lines).

The data and model predictions addressing the main hypothesis of this simulation are in the lower row of panel B. The central training data show a large positive deviation in the middle of the target range, corresponding to the mixed reference frame, while the peripheral training data are always close to zero, an evidence of the head-centered RF. The HEC model (magenta line) approximates this pattern by predicting a positive deviation in both training regions accompanied by a negative deviation of similar size for the targets to the left of the training regions. This pattern captures the main characteristics of the data even though the predicted positive deviation is weaker than that observed for the central central-training data. On the other hand, the HC model (brown line) always predicts no deviation from zero, as that model assumes that the adaptation always occurs in the HC RF. These differences between the models confirm the hypothesis that auditory representation is adapted uniformly by visual signals in both head-center and eye-center reference frames. This conclusion is confirmed by the AICc evaluation which showed almost no support for the HC model compared to the HEC model ($\Delta \mathrm{AIC} = 7.9$).

The model predictions for the no-shift data (upper row of panel B) are almost identical for the two models. Thus, the difference in performance between the models cannot be explained by differences in accounting for the no-shift data. Notably, the predictions for the two training regions are fairly similar to each other, and slightly worse than those obtained in the No Shift simulation. However, they still capture the pattern of biases fairly well. Finally, note that the predictions for the average of positive and negative shift data is not shown, even though these transformed data were also used for fitting. These data were omitted as both the data and model predictions are very similar to the no-shift results shown in the upper row of panel B.

Looking at across-subject variability in the data, the error bars in panel B tend to be smaller for the positive-vs-negative shift plots (lower row) than for the no-shift plots (upper row). This difference is in fact much larger, since the plotted error bars are for the difference between the two shift directions, whereas the aftereffect magnitude equal to half of the difference was used in the fitting. This shows that additional between-subject variability was caused by idiosyncratic biases in each subject's responses that are consistent within each subject, and which therefore cancel out when the difference between positive and negative shift data is computed. This again shows the importance of fitting the models on the transformed data, which resulted in weighing the positive-vs-negative shift difference data (lower row) even more than the no-shift training-vs-non-training FP data (upper row).

Panel A illustrates the behavior of individual components that resulted in the models' predictions. The saccade-related bias is almost identical for the two models (upper row), and overall similar to the pattern observed in the NoShift simulation (Figure 23A). The auditory space adaptation is broad for both models, and only slightly different between the models (magenta vs. brown lines between in the lower row of Figure 24B). The size of the difference is mainly determined by parameter w_E (see Table 3) which defines the relative contribution of the eye-centered vs. head-centered RF to the combined representation in the HEC model (in this simulation $w_E = 0.15$, indicating that the EC RF only had a 15% weight in the mixed reference frame). So, it can be concluded that even though this contribution is highly significant, the HC RF has still a dominant role when uniform representation of the auditory space is assumed.

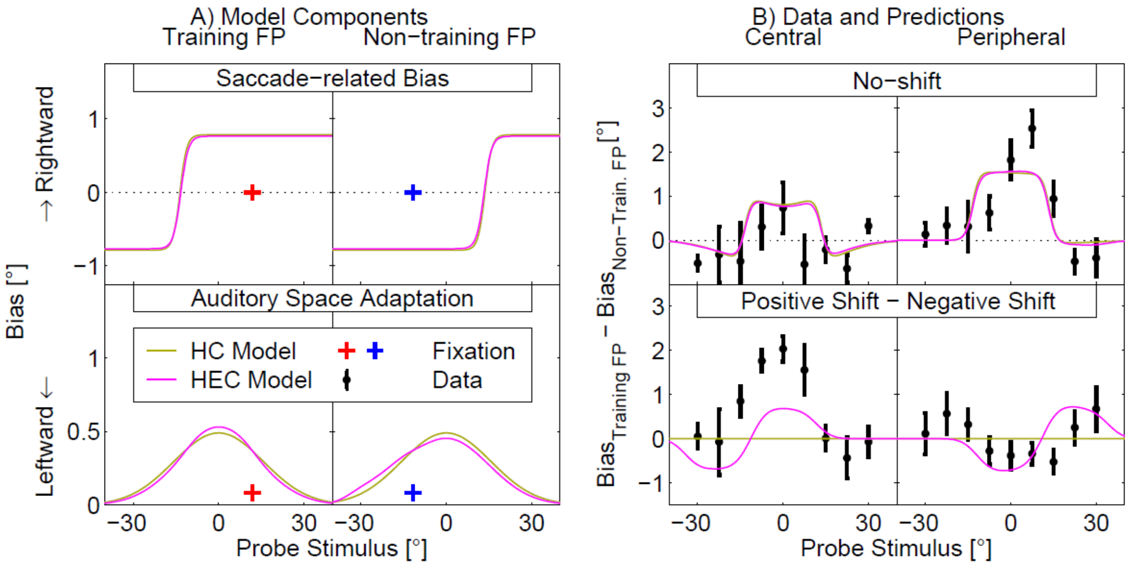

Figure 24: Model predictions and data for the All Data simulation. A) Visualization of the two model components, Saccade-Related Bias and Auditory Space Adaptation, for the HC and HEC models with the parameters fitted to all the data (from Table 3). For detailed description see the caption for panel A of Fig. 3. B) Across-subject mean difference in biases from the training FP vs. non-training FP (±standard error of the mean) and model predictions for the no-shift data, and for the aftereffect magnitude computed as a difference between positive-shift and negative-shift data (lower row).

6.6.3 Central and Peripheral Data simulations

Two additional simulations were performed, each of them fitting separately the data for only one training region. The main goal of the simulations was to verify that, when the models are fitted to the two data sets separately, they will confirm the conclusions of the behavioral experiments about the mixed reference frame for the central-training data and the head-centered reference frame for the peripheral-training data. Additionally, these simulations only fitted the transformed positive-shift and negative-shift data, while also producing model predictions for the no-shift data. Thereby, the simulations tested whether the behavior of the saccade-related model component observed in the previous simulations is dependent on the presence of the no-shift data, or whether the models would find a similar predicted pattern even if only the positive/negative shift data are considered.

Central Data simulation fitted only the central-training data from the positive-shift and negative-shift conditions (dashed lines in Figure 21B). The main hypothesis tested in the simulation was that the *RF is mixed when VAE is induced in the central region*. This hypothesis would be confirmed if the HEC model is significantly better than the HC model. Figure 25 presents the results of this simulation using a layout identical to Figure 24. The lower row of panel B shows the predictions of the two models for the difference data. As expected, the HEC model (magenta) fits the central-training data well (better than in the All Data simulation) while the HC model's prediction (brown) is again fixed at zero. This difference confirms the hypothesis that the EC RF contributes significantly to the ventriloquism adaptation in central region, a conclusion also confirmed by the AICc evaluation (HEC model better than HC model; $\Delta\mathrm{AIC} = 5.9$). However, it is also noticeable that the HEC model underestimates the central data for targets at azimuths around 0° while it predicts a negative deviation at azimuths around -20°, not observed in the data. This negative deviation is due to the structure of the model which always predicts that a positive deviation is accompanied by a negative deviation at locations shifted in the direction of the new, non-training FP location. For the peripheral experiment, the HEC model predictions depart considerably from the data, as expected since the data do not show a strong EC RF contribution. On the other hand, for the no-shift data, both models largely capture the main trends even though they were not fitted on these data (upper row of panel B), confirming that the FP-dependent adaptation observed in the no-shift data is not specific to these data as the model generalizes to predict it even if only trained on the positive and negative shift data.

Considering the individual model components (Panel A), the results are overall similar to the All Data simulation (Figure 24). The main difference in the current simulation is that the EC-referenced contribution to auditory spatial adaptation in the HEC model is considerably stronger, resulting in larger differences between the two models (bottom row). However, even here the HC RF still has more weight ($w_E = 0.3$ in Table 3), suggesting that it is the more dominant RF for ventriloquism aftereffect in general.

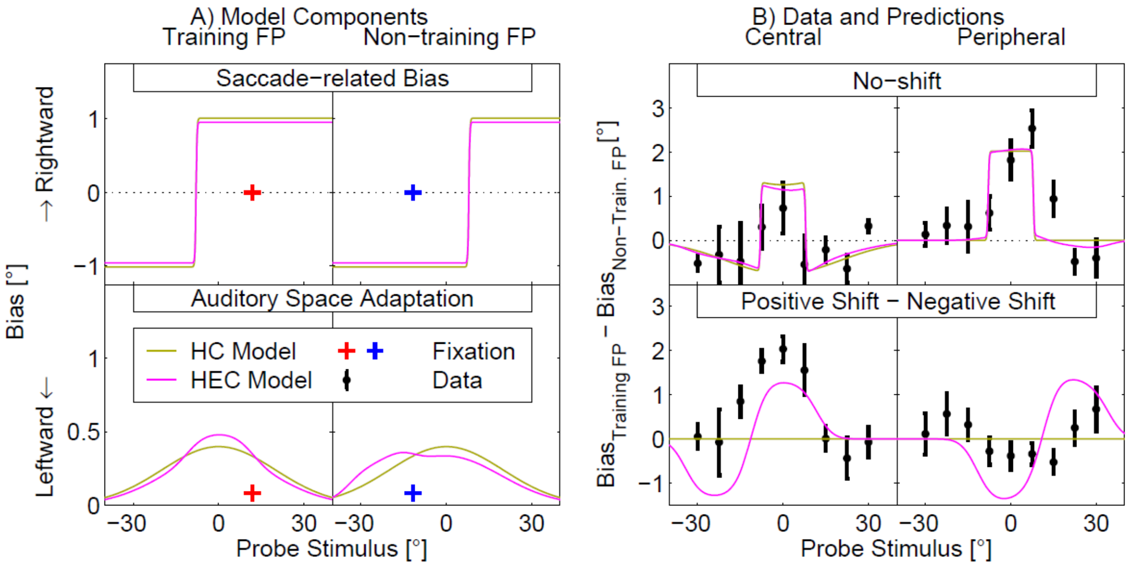

Figure 25: Model predictions and data for the Central Data simulation. For detailed description, see the caption of Figure 24.

Peripheral Data simulation fitted only the peripheral-training data from the positive-shift and negative-shift conditions (dashed lines in Figure 21B). The main goal was to confirm the hypothesis that the *RF is head-centered when VEA is induced in the peripheral region*, in agreement with the behavioral results. This hypothesis would be confirmed if the HEC and HC models performed similarly in the simulation.

Figure 26 presents the results of this simulation using a layout identical to Figure 24. The lower row of panel B shows the predictions of the two models for the positive vs. negative shift difference data. As expected, both models fit the near-zero peripheral-training data well, while failing to predict the central-training data. This confirms that the EC RF does not contribute to the ventriloquism adaptation in the peripheral region, a conclusion also supported by the AICc evaluation, in which the HC model is better than the HEC model; $\Delta\text{AIC} = 5.6$ in Table 3). Similar to the Central Data simulation, for the no-shift data, both models largely captured the main trends even though they were not fitted on these data (upper row of panel B). These results are also confirmed when considering the individual model components (Panel A). First, the saccade-related bias component (upper row) again behaves identically in the two models similarly to the previous simulations. Second, the auditory space adaptation component (lower row) behaves nearly identically for the two models, determined by the low the relative weight of the EC RF in the HEC model ($w_E = 0.04$ in Table 3).

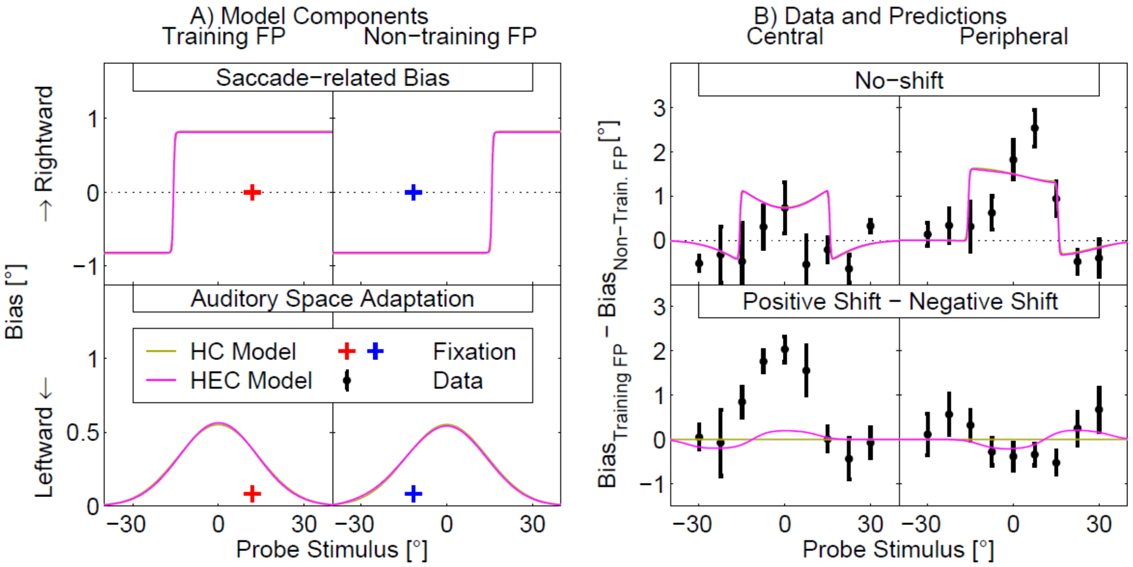

Figure 26: Model predictions and data for the Peripheral Data simulation. For detailed description, see the caption of Figure 24.

6.6.4 Model parameter values

The behavior of the models in different conditions can be analyzed by looking at the fitted values of the model parameters. Here, the first main modeling question concerned the ability of the models to predict the EC-dependence of the no-shift data observed in the peripheral, but not in the central, training condition. The critical model parameters here are the parameters h and w, which determine the relative strength of the saccade-related and auditory space adaptation components of the model (Figure 27 and Table 3). The values of the two parameters are overall similar in all simulations, suggesting that both components contributed critically to all the predictions.

The parameter w_E determined the relative strength of the EC RF contribution to the ventriloquism-driven auditory spatial adaptation, while the parameters σ_H and σ_E determined, respectively, how broad-vs-specific was the influence of the HC and EC RFs. The value of w_E was always much smaller than 0.5 (in relevant simulations smaller or equal to 0.3) and σ_H was always much larger than σ_E. Both these observations indicate that while the EC-referenced signals influence the ventriloquism adaptation significantly, their effect is mostly modulatory, while the HC-referenced signals dominate.

Finally, the fitted values of parameters k and c did not change dramatically across the simulations, always resulting in similar predictions about the saccade-related bias component of the model.

6.7 Summary and Discussion

The HC/HEC model introduced here aims to characterize the reference frame in which auditory and visual signals are combined to induce the ventriloquism aftereffect. It focuses on the experimental data in which ventriloquism was induced locally in either the audiovisual center or periphery, in which a change in fixation point was used to dissociate the head-center from eye-centered reference frames, and in which saccades were used for responding during training and testing (Kopco *et al.*, 2009; Kopco *et al.*, 2019b). The model assumes a population of adaptive units representing the auditory space with auditory and visual inputs, similar to the channel processing model proposed in (Carlile *et al.*, 2001). However, instead of explicitly implementing a population of units, it describes the adaptive effects by only considering the locations from which the auditory components of audiovisual training stimuli were presented. Then, for each unit there is a Gaussian neighborhood in which the AV training affects the A-only responses in either HC-only RF (HC model) or in a combined HC+EC RF (HEC model). Also, the model assumes that there are intrinsic biases associated with auditory saccade responses, and that the effect of ventriloquism is to shift the auditory-only responses from these saccade-related biases towards the locations of the responses on the audiovisual training trials.

Since the model only uses the responses on audiovisual training trials to guide adaptation, independent of the direction of audiovisual disparity used during training, and independent of whether the adaptation results in hypometric or hypermetric saccades, it is assumed that there is a direct relation between the audiovisual responses during training and the auditory-only responses during testing. Specifically, the assumed relationship is that the ratio of observed ventriloquism aftereffect to the observed ventriloquism effect is constant, as confirmed by our behavioral data analysis (see Appendix) which found a ratio of approximately 0.5. This ratio is not aftereffect by whether the aftereffect results in hypometric or hypermetric saccades, consistent with Pages and Groh (2013). However, the analysis also found that there is an asymmetry in the ventriloquism effect when measured using audiovisual saccades. Specifically, the effect reaches 100% of audio-visual disparity if resulting in hypometric saccades, while it is only 80% of the disparity when resulting in hypermetric saccades. Future studies will need to determine whether there is really a difference in the presence/absence of the hypo/hypermetric asymmetry when saccades are used for ventriloquism effect and aftereffect measurement, or whether the current results are different for the effect vs. aftereffect only because the aftereffect data are noisier.

The four simulations presented here showed that the HC/HEC model can describe the different phenomena observed in the Kopco et al. (Kopco *et al.*, 2009; Kopco *et al.*, 2019b) studies. First, in the No-Shift simulation, the simpler HC model accurately predicted the newly reported adaptation by AV-aligned stimuli (Kopco *et al.*, 2019b) as a combination of the intrinsically present saccade-related biases locally "corrected" by the visually guided adaptation at the training locations. Thus, the model predicts that this AV-aligned adaptation for the peripheral-training data is purely driven by some adaptive processes affecting the motor representations related to audiovisual/auditory saccades. This, as well as the existence of the saccade-related bias component of the model, can be tested in future studies, as the currently available data are not consistent as to whether auditory saccades are predominantly hypermetric or hypometric (Yao and Peck, 1997; Gabriel *et al.*, 2010). Both these predictions can be experimentally tested by performing ventriloquism experiments in which saccades are not used for responding (Kopco *et al.*, 2015).

The second, All Data simulation addressed the main question of this study about the reference frame of the ventriloquism aftereffect. Its results provide an evidence that a uniform auditory spatial representation uses a mixed reference frame, with visual signals adapting the auditory spatial representation in both head-centered and eye-centered RFs, as implemented in the HEC model and consistent with physiological studies (Mullette-Gillman *et al.*, 2005; Porter and Groh, 2006). Importantly, the current results suggest that, in the mixed frame, the relative contribution of the EC RF is only 15% vs. 85% for the HC RF. Moreover, even when only the central-training data are considered (Central-Data simulation), the relative contribution of the EC only reaches 30%. Thus, the HC RF is always dominant for the ventriloquism aftereffect adaptation, an observation that is further supporter by the comparison of the fitted sigma parameters (which showed that the HC-referenced adaptation is more broad than the EC-referenced adaptation). The second simulation also showed that the model in its current form always predicts the same difference in biases between the FPs, independent of the training region. This effect is mainly due to the implicit model assumption that the distribution of the spatial channels is uniform across space. If the model assumed a denser representation of space near the midline (e.g., see (Stern and Shear, 1996)), it could predict adaptation that is stronger in the center than in the periphery.

Importantly, the current model was fitted on data transformed so that the differences between the two FPs and differences between the positive and negative shift data were used. This was particularly critical for this simulation in which the EC contribution is visible when the double difference is computed, and it was also important since, in this representation, a lot of noise in the data is removed. Note that when the All data simulation was repeated on untransformed data, the AICc evaluation did not find a significant difference between the HC and HEC models, since the across-subject variability in the responses considered separately for the two FPs was too large, dominating over the differences between the FPs critical to evaluate the reference frames (data not shown).

The final two simulations examined the model behavior when fitted separately to the central vs. peripheral training data. In both simulations the model predictions were in agreement with the behavioral data. Specifically, the HEC model using a mixed reference frame better predicted the central data, while the HC model using the head-centered reference frame better predicted the peripheral data. The central-data simulation also showed one weakness of the model: in its current form it always predicts that if there is a region in which VAE magnitude is larger for the training-FP than non-training-FP data, then there also has to be a region in which the relationship is reversed. An extension of the model which would make the strength of the adaptation depend not only on the distance from the training stimuli, but also on the distance from the training FP, could correct this discrepancy.

Finally, the Central and Peripheral Data simulations accurately captured the no-shift data, even though the models were not fitted on them, confirming that the pattern of adaptation exhibited in these data is also present in the positive-shift and negative-shift data from which it can generalize to the no-shift data. However, as discussed above, the no-shift data biases are most likely related to the saccade responses, not to the spatial representation adapted by ventriloquism, which is of primary interest here.

The neural mechanisms of the ventriloquism aftereffect and its reference frame are not well understood. Cortical areas involved in ventriloquism aftereffect likely include Heschl's gyrus, planum temporale, intraparietal sulcus, and inferior parietal lobule (Zatorre *et al.*, 2002; Michalka *et al.*, 2016; Zierul *et al.*, 2017; van der Heijden *et al.*, 2019). Multiple studies found some form of hybrid representation or mixed auditory and visual signals in several areas of the auditory pathway, including the inferior colliculus (Zwiers *et al.*, 2004), primary auditory cortex (Werner-Reiss *et al.*, 2003), the posterior parietal cortex (Duhamel *et al.*, 1997; Mullette-Gillman *et al.*, 2005; 2009), as well as in the areas responsible for planning saccades in the superior colliculus and the frontal eye fields (Schiller *et al.*, 1979; Wallace and Stein, 1994). In the current model, the saccade-related component likely corresponds to the saccade-planning areas. The auditory space representation component likely corresponds to the higher auditory cortical areas or the posterior parietal areas, not the primary cortical areas. This can be expected because there is growing evidence that, in mammals, auditory space is primarily encoded non-homogeneously, based on two spatial channels roughly aligned with the left and right hemifields of the horizontal plane (McAlpine *et al.*, 2001; Stecker *et al.*, 2005; Salminen *et al.*, 2009; Grothe *et al.*, 2010; Groh, 2014) and the ventriloquism adaptations modeled here are local (within a hemifield or just in the central region), not consistent with broad adaptation predicted by the hemifield code. However, note that there are also theories which incorporate additional channels, such as a central channel, in addition to the hemifield channels (Dingle *et al.*, 2012). Such extended models might be compatible with the current data.

Even though most previous recalibration studies examined the aftereffect on the time scales of minutes (Radeau and Bertelson, 1974; 1976; Recanzone, 1998b; Woods and Recanzone, 2004), recent studies demonstrated that it be elicited very rapidly, e.g., by a single trial with audio-visual conflict (Wozny and Shams, 2011). If it is the case that the adaptive processes underlying the ventriloquism aftereffect occur on multiple time scales, as also suggested in several models of slower ventriloquism aftereffect (Bosen *et al.*, 2018; Watson *et al.*, 2019), then an open question is whether the reference frame is the same at the different scales or whether it is different. The current results are mostly applicable to the slow adaptation on the time scale of minutes, while the RF on the shorter time scales has not been previously explored.

In summary, while some previous models considered the reference frame of the ventriloquism effect (Pouget *et al.*, 2002; Razavi *et al.*, 2007), the current HC/HEC model is, to our knowledge, the first one to focus on the RF of the ventriloquism aftereffect. In addition, it also considers how saccade-related adaptation might influence auditory saccades. In the future, it can be combined with the existing models of spatial and temporal characteristics of the ventriloquism aftereffect to obtain a more general model of this important multisensory phenomenon.

Acknowledgments

This work was supported by VEGA-1/0355/20 and VVGS UPJS VVGS-2020-1514.

6.8 Appendix

To examine whether auditory saccades used for responding have properties that might be important for the current modeling, responses to auditory and audiovisual stimuli in the training regions of both experiments were further analyzed (Figure 27). Two questions were addressed. First, we examined whether the observed saccades were longer or shorter depending on whether the presence of visual component/adaptation resulted in saccades that were hypometric (shorter than needed to reach the auditory target) or hypermetric (longer than needed to reach the auditory target). Such asymmetry, if observed, would suggest that some of the effects described in Section 2, e.g., the eye-centered RF effects, might have been caused by the saccade responses. Second, we evaluated whether the ratio of the magnitudes in auditory-only responses to the respective AV responses for a given AV stimulus is constant for all combinations of audiovisual stimuli. If that is the case, then, independent of any possible hypo/hypermetric dependence, the model can assume that the predicted ventriloquism aftereffect is directly related to the measured ventriloquism effect.

Figure 27A shows the biases in saccade responses from the training FP for targets in the training regions from both experiments (circles vs. squares). Open symbols represent audio-visual responses, filled symbols auditory-only responses. Black symbols represent the AV-aligned runs, while the cyan and magenta symbols represent, respectively, the runs in which the response shifts towards the visual component/adaptation resulted in saccades that were hypometric and hypermetric. Specifically, the magenta circles represent the central-training data with visual component shifted to the right, i.e., towards the fixation point, while the magenta squares represent the peripheral-training data with visual component shifted to the left, i.e., again towards the fixation point (the cyan data then represent the corresponding data for visual components shifted in the opposite direction). Note that the filled symbols here show the same data as the red lines in the training regions of Figure 21B, C.

The black symbols in Figure 27A show that, in both experiments, all the saccades in the AV-aligned runs were fairly accurate. Specifically, responses to the AV stimuli were within +/-0.5° (open black symbols) while the saccades to the auditory targets (filled black symbols) tended to be hypometric (rightward bias for targets to the left of the FP and leftward for the targets to the right) by up to 1°, except for one data point (7.5°), discussed in detail later.

Comparison of the respective magenta and cyan symbols shows that the adaptation direction (i.e., visual component displacement) that led to hypometric saccades tended to result in larger biases than the direction leading to hypermetric saccades (for example, all the magenta filled circles are clustered around the value of 3, while the corresponding cyan filled circles are around -1). To analyze this asymmetry while accounting for the biases in the AV-aligned responses, Figure 27B shows the hypometric and hypermetric data from panel A referenced to the respective baselines and plotted such that positive values always represent bias in the direction of the visual component displacement (i.e., all the cyan squares and magenta squares had their signs flipped after subtracting the baseline). The magenta open symbols show that, independent of the training region, the VE responses measured in conditions resulting in hypometric saccades were aligned with the visual component (which was separated by 5°), while the responses resulting in hypermetric saccades (open cyan symbols) only reach approximately 80% of the visual component displacement. A mixed ANOVA with a between-subject factor of Experiment (Central, Peripheral) and within-subject factors of Shift Direction (Hypometric, Hypermetric), and Azimuth (Small, Medium, Large) performed on these data confirmed these results, showing a significant main effect of shift direction ($F(1,12) = 5.78$; $p = 0.033$). The ANOVA also found a significant Azimuth X Experiment interaction ($F(2,24) = 9.71$; $p = 0.006$) reflecting a dependence of the effect on the target location that is not further considered here, and no other significant main effects or interactions ($p > 0.1$). On the other hand, for the VAE data, no significant difference between hypometric and hypermetric saccades was observed (a similar ANOVA on these data only found a main effect of Azimuth; $F(2,24) = 7.94$; $p = 0.002$). Thus, the strong asymmetry between the hypometric and hypermetric AV data in in panel A (filled cyan vs. magenta symbols) can be ascribed to overall hypometry of the auditory saccades exhibited also by the No-Shift data (black filled symbols). Also note that there is one hypermetric AV data point for which the response referenced to baseline is near 0 (left-most filled cyan circle), not following the pattern observed for all the other points. Most likely, this inconsistency is caused by some specific characteristic of the baseline auditory-only saccades, as this point corresponds to the only black filled symbol that shows hypermetry instead of hypometry in panel A (the black filled circle at the 7.5° location).

Finally, panel C shows the observed VAE as a proportion of the observed VE (i.e., each symbol in panel C shows the ratio of the corresponding filled and open symbols from panel B). In this analysis, one subject was identified as outlier (in at least one data point the subject differed from the across-subject mean by more than 3 standard deviations). This subject is plotted separately (crosses) and not included in the across-subject graphs. For the remaining subjects, Figure 27C shows that there is a constant relationship between the induced ventriloquism effects and aftereffects such that the aftereffect is always approximately one half of the effect (with a slight tendency to grow with the target amplitude), independent of whether the shift is hypo/hypermetric or of the training region. Confirming this observation, ANOVA with the same factors as above only found a main effect of Azimuth (F(2,22)=10.34, p=0.0007). The only other factor that approached significance was Training Region (F(1, 11)=3.83, p=0.076) while all the other factors and interactions were not significant (p > 0.15). These results are used in the current modeling in which it is assumed that there is a constant relationship between the induced ventriloquism effect and aftereffect, independent of whether the induced shift is hypometric or hypermetric.

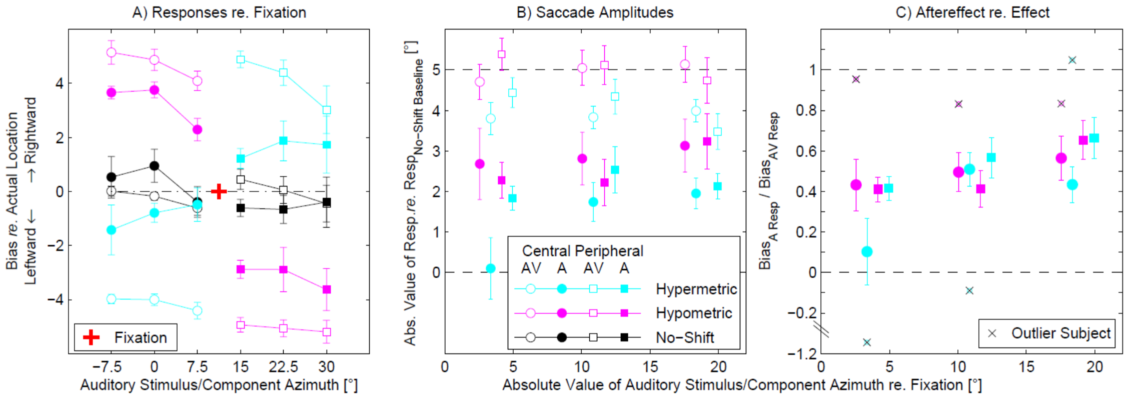

Figure 27: Saccade responses to audiovisual and auditory stimuli in the training regions from both experiments. A) Across-subject mean saccade end points as a function of the location of the auditory target or of the auditory component of the audio-visual target. Data are plotted separately for the auditory and audio-visual stimuli, for the two training regions, and for the three directions of the visual component displacement (aligned, shifting the auditory saccade to be hypometric, shifting the auditory saccade to be hypermetric). Note that a hypometric shift corresponds to visual component shifted to the right for the central-training data and to visual component shifted to the left for the peripheral training data (and vice versa for the hypermetric shift). B) Strength of the induced ventriloquism effect and aftereffect shown as the across-subject mean bias in response towards the visual component re. response in no-shift baseline (i.e., difference between the respective magenta/cyan and black symbols from panel A with the sign flipped for the negative-shift data). C) Ventriloquism aftereffect as a proportion of ventriloquism effect shown as the across-subject mean ratio of the VE/VAE strengths from panel B. Note that one outlier subject is plotted separately from the across-subject means in this analysis. Error bars represent across-subject standard errors of the means (N=7 in both experiments).

Conclusion

This thesis examined the reference frame of the ventriloquism aftereffect by analyzing behavioral data and developing several computational models. The first main result of the behavioral analysis is that the reference frames are inconsistent for the central (mixed reference frame) vs. peripheral (head-centered reference frame) inducement of the ventriloquism aftereffect. The second main result is that the ventrilqousim aftereffect differs for the central vs. peripheral also in the case that it was induced by the audiovisual stimuli with the visual vs. auditory components aligned.

And these two insonsistencies were together the reason to perform the modeling. In the first preliminary modeling the additivity of the effect of training region and the effect of eye-gaze was assumed. Such assumption was proven wrong. In the second preliminary modeling paper, a more complex model was proposed which was also shown as perspective for AV-aligned data, but proven wrong for the AV-misaligned data. The main modeling is then presented. The main conclusion of this modeling is that according to the modeling and its results the reference frame of the ventriloquism aftereffect is basically head-centered with significant eye-centered modulation.

These results are important for the basic understanding of the neural processes underlying cross-modal integration in the human brain. But, they can also be useful in the design of audio-visual virtual reality systems or for systems for trainting auditory spatial skills in normal-hearing and hearing-impaired listeners.

Resumé

Predstavili sme jednu analytickú a tri modelovacie štúdie vzťažnej sústavy bruchomluveckého afterefektu, pričom všetky štyri sú rozšírením prevej štúdie na túto tému (Kopčo a spol., 2009).

Bruchomluvecký afterefekt (VA z angl. ventriloquism aftereffect) spočíva v kalibrácií priestorového sluchu v priestore lokalizovateľnými zvukmi synchrónne sprevádzanými zrakovými stimulmi prehrávanými z blízkeho ale odlišného miesta v priestore. Takáto kalibrácia sa prejavuje následným posunom v lokalizácií už žiadným iným stimulom nesprevádzanými zvukmi.

Mozog používa rôzne metódy na zakódovanie pozícií zrakových verzus sluchových podnetov. Sietnica sníma pozície zrakových objektov s ohľadom na oči, ktoré sa hýbu, keď sa obzeráme. Sluchový systém získava priestorovú informáciu z rozdielov v charakteristikách zvuku pre jedno verzus druhé ucho, čo ukazuje, že pozície zdrojov zvuku sú vztiahnuté na hlavu. A je potrebné, aby tieto dve vzťažné sústavy boli zarovnané, aby sa vytvoril koherentny sluchovo-zrakový priestorový vnem, alebo aby sa mohla uskutočniť zrakom riadená rekalibrácia sluchového priestoru.

Cieľom tejto práce je preskúmať a modelovať procesy zarovnania súradnicových sústav reprezentácie zraku a sluchu v mozgu. Známym javom, v ktorom zrak a sluch interagujú, je práve bruchomluvecký afterefekt.

Štúdia (Kopčo a spol., 2009) ukázala že vzťažná sústava bruchomluveckého afterefektu je zmesou vzťažných sústav orientovaných na polohu hlavy (HC – angl. head-centered) a polohu očí (EC – angl. eye-centered), kde VA je vyvolaný v strede sluchovo-zrakového poľa.

Čo sa týka experimentalnych metód, v našej analytickej štúdií a aj v štúdií Kopčo a spol. (2009) bol vyvolávaný VA vždy tak, aby sa experimentálny subject pri prezentácií každého audiovizuálneho tréningového stimulu pozeral na ten istý fixačný bod, čím sa zabezpečilo, aby bol VA vyvolaný na tom istom mieste podľa oboch vzťažných sústav (aj HC, aj EC), vzhľadom na to, že poloha hlavy je u subjektov počas experimentu vždy zafixovaná. No ďalej sa pre meranie VA, teda pri čisto sluchových stimuloch, používali až dva fixačné body, aby bola identifikovaná vzťažná sústava VA.

Experimenty pozostávali z väčšieho množstva sedení, ktoré sa líšili okrem miesta vyvolávania VA aj smerom posunu zrakového adaptora oproti synchrónnemu zvuku. A boli použité tieto tri nastavenia: pozitívne, negatívne a zarovnané. No my sme naše dáta spracovávali tak, že namiesto dát pre pozitívne a negatívne posuny sme sa pozerali hlavne na ich rozdiel, ktorý sme po veľmi jednoduchej úprave nazývali aj veľkosťou afterefektu.

Najdôležitejšou premennou, ktorá bola meraná pri každej jednej iterácií experimentálneho sedenia, bola sakadická odpoveď očí subjektu, ktorý dostal pokyn, aby sa pri každej jednej iterácií pozrel na miesto, odkiaľ počul prichádzať zvuk. Táto premenná bola pre ciele našich analýz a modelovania vo väčšine prípadov prekonvertovávaná na orientovanú odchýlku odpovede od pozície prichádzajúceho zvuku (skrátene: odchýlku).

V našej analytickej štúdií boli analyzované výsledky nového experimentu, v ktorom bol tento efekt vytvorený na rozdiel od pôvodnej štúdie v sluchovo-zrakovej periférii, aby sa preskúmalo, či uniformity v kódovaní audiovizuálneho priestoru ovplyvňujú pozorovanú vzťažnú sústavu. V periférií bola vzťažná sústava identifikovaná ako primárne centrovaná na polohu hlavy. Výsledky tiež ukázali novú formu zrakom spôsobenej adaptácie v sluchovej reprezentácií, ktorá závisela na informáciách centrovaných na polohu aj hlavy aj očí.

Vzhľadom na nekonzistentnosť výsledkov našej štúdie so štúdiou Kopčo a spol. (2009) bolo vykonané modelovanie, a čo sa týka našich modelovacích štúdií, prvé dve boli predbežné a tretia bola hlavná. V každej z nich boli navrhnuté, fitované a testované rôzne verzie modelu.

Finálny model mal dve hlavné verzie, z ktorých každá obsahovala dve aditívne skombinovné zložky: sakadická (týkajúca sa pohybov očí) zložka charakterizujúca adaptáciu sluchových sakadických odpovedí, a zložka pre priestorovo-sluchovú reprezentáciu adaptovanú bruchomluveckými signálmi.

Sakadická zložka závisela iba od polohy fixačného bodu a miesta prichádzajúceho zvuku, pričom ako funkcia miesta pricházajúceho zvuku mala tvar tromi parametrami škálovanej sigmoidy.

Zložka pre priestorovo-sluchovú reprezentáciu závisela okrem fixačného bodu a miesta prichádzajúceho zvuku aj od tréningových audiovizuálnych stimulov a odpovedí subjektu na ne.

A teda obe verzie modelu sa odlišovali v tom, či boli signály vo vzťažnej sústave centrované na polohu hlavy (HC verzia), alebo boli kombináciou vzťažných sústav centrovaných na polohy hlavy a očí (HEC verzia).

Tieto obe verzie boli evaluované v štyroch simuláciách, pričom tieto simulácie sa odlišovali v tom, ktorá podmnožina našich experimentálnych dát bola zohľadnená. HEC model mal lepšie výsledky v porovnaní s modelom HC v hlavnej simulácií, ktorá brala do úvahy všetky dáta, kým HC model bol vhodný keď bola adaptácia vyvolávaná len zarovnanými sluchovo-zrakovými stimulmi.

Tieto výsledky podporujú závery, že kým je VA ovplyvňovaný viacerými priestorovými nerovnomernými hemisfericky-špecfickými procesmi, na kalibráciu sluchovej priestorovej reprezentácie sú použité vizuálne signály v uniformnej zmiešanej HC+EC vzťažnej sústave, dokonca aj po zohľadnení EC sluchovej sakadickej adaptácie. Teda, vplyv EC je prítomný v obidvoch zložkách modelu, zložky sakadickej a aj zložky pre priestorovo-sluchovú reprezentáciu adaptovanú bruchomluveckými signálmi.

References

Ahveninen, J., Kopco, N., and Jaaskelainen, I. P. (**2014**). "Psychophysics and neuronal bases of sound localization in humans," Hearing Res **307**, 86-97.

Alais, D., and Burr, D. (**2004**). "The ventriloquist effect results from near-optimal bimodal integration," Curr Biol **14**, 257-262.

Bertelson, P., Frissen, I., Vroomen, J., and de Gelder, B. (**2006**). "The aftereffects of ventriloquism: Patterns of spatial generalization," Percept. Psychophys. **68**, 428-436.

Bosen, A. K., Fleming, J. T., Allen, P. D., O'Neil, W. E., and Paige, G. D. (**2018**). "Multiple time scales of the ventriloquism aftereffect," Plos One **13**.

Brainard, M. S., and Knudsen, E. I. (**1995**). "Dynamics of visually guided auditory plasticity in the optic tectum of the barn owl," J. Neurophys. **73**, 595-614.

Brown, G. J., Beeston, A. V., and Palomaki, K. J. (**2012**). "Perceptual compensation for the effects of reverberation on consonant identification: A comparison of human and machine performance," in *13th Annual Conference of the International-Speech-Communication-Association* (Portland, OR), pp. 1714-1717.

Bruns, P., Dinse, H. R., and Roder, B. (**2020**). "Differential effects of the temporal and spatial distribution of audiovisual stimuli on cross-modal spatial recalibration," Eur J Neurosci **52**, 3763-3775.

Bulkin, D., and Groh, J. M. (**2012**). "Distribution of eye position information in the monkey inferior colliculus," Journal of Neurophsyiology **107**, 785-795.

Burnham, K. P., and Anderson, D. R. (**2004**). "Multimodel Inference Understanding AIC and BIC in Model Selection," Sociological Methods & Research **33**, 261-304.

Calvert, G. A., Spence, C., and Stein, B. E. (**2004**). *The Handbook of Multisensory Processes*.

Canon, L. K. (**1970**). "Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation," J. Exp. Psych. **84**, 141-147.

Carlile, S., Hyams, S., and Delaney, S. (**2001**). "Systematic distortions of auditory space perception following prolonged exposure to broadband noise," J. Acoust. Soc. Am. **110**, 416-424.

Dahmen, J. C., Keating, P., Nodal, F. R., Schulz, A. L., and King, A. J. (**2010**). "Adaptation to Stimulus Statistics in the Perception and Neural Representation of Auditory Space," Neuron **66**, 937-948.

Dingle, R. N., Hall, S. E., and Phillips, D. P. (**2012**). "The three-channel model of sound localization mechanisms: interaural level differences," J Acoust Soc Am **131**, 4023-4029.

Duhamel, J.-R., Bremmer, F., BenHamed, S., and Graf, W. (**1997**). "Spatial invariance of visual receptive fields in parietal cortex neurons," Nature **389**, 845-848.

Feddersen, W. E., Sandel, T. T., Teas, D. C., and Jeffress, L. A. (**1957**). "Localization of high-frequency tones," J. Acoust. Soc. Am. **29**, 988-991.

Gabriel, D. N., Munoz, D. P., and Boehnke, S. E. (**2010**). "The eccentricity effect for auditory saccadic reaction times is independent of target frequency," Hearing Res **262**, 19-25.

Groh, J. M. (**2014**). "Making space: how the brain knows where things are.," Cambridge, MA: Harvard University Press.

Groh, J. M., and Sparks, D. L. (**1992**). "Two models for transforming auditory signals from head-centered to eye-centered coordinates," Biological Cybernetics **67**, 291-302.

Grothe, B., Pecka, M., and McAlpine, D. (**2010**). "Mechanisms of sound localization in mammals.," Physiol Rev **90**, 983-1012.

Haessly, A., Sirosh, J., and Miikkulainen, R. (**1995**). "A model of visually guided plasticity of the auditory spatial map in the barn owl," in *Seventeenth Annual Meetings of the Cognitive Science Society* (Erlbaum, Pittsburgh, PA), pp. 154-158.

Harris, C. M. (**1995**). "Does saccadic undershoot minimize saccadic flight-time? A Monte-Carlo study," Vision Res **35**, 1995.

Hendrickx, E., Paquier, M., and Palacino, J. (**2015**). "Ventriloquism effect with sound stimuli varying in both azimuth and elevation," J Acoust Soc Am **138**.

Howard, I. P., and Templeton, W. B. (**1966**). *Human Spatial Orientation* (Wiley, London).

Choe, C. S., Welch, R. B., Gilford, R. M., and Juola, J. F. (**1975**). "The ‘ventriloquism effect: ’ visual dominance or response bias?," Percept. Psychophys. **18**, 55-60.

Jack, C. E., and Thurlow, W. R. (**1973**). "Effects of degree of visual association and angle of displacement on the ‘ventriloquism’ effect," Percept. Mot. Skills **37**, 967-979.

Johnson, H. M. (**1920**). "The Dynamogenic Influence of Light on Tactile Discrimination," Psychobiology **2**, 351.

Klatzky, R. L., and Lederman, S. J. (**2010**). "Multisensory texture perception.," in *Multisensory object perception in the primate brain*, edited by M. J. Naumer, and J. Kaiser (Springer Science + Business Media), pp. 211-230.

Knudsen, E. I., and Knudsen, P. F. (**1985**). "Vision guides the adjustment of auditory localization in young barn owls," Science **230**, 545-548.

Knudsen, E. I., and Knudsen, P. F. (**1989**). "Vision calibrates sound localization in developing barn owls," J. Neurosci. **9**, 3306-3313.

Kopco, N. (**2020**). "Cross-modal interactions," Introduction to Neuroscience Lectures **13**, 4-25.

Kopco, N., Lin, I. F., Shinn-Cunningham, B. G., and Groh, J. M. (**2009**). "Reference Frame of the Ventriloquism Aftereffect," J. Neurosci. **29**, 13809-13814.

Kopco, N., Loksa, P., Lin, I. F., Groh, J., and Shinn-Cunningham, B. (**2019a**). "Hemisphere-Specific Properties of the Ventriloquism Aftereffect in Humans and Monkeys," (www.biorxiv.org).

Kopco, N., Loksa, P., Lin, I. F., Groh, J., and Shinn-Cunningham, B. (**2019b**). "Hemisphere-specific properties of the vnetriloquism aftereffect," J Acoust Soc Am **146**, EL177-183.

Kopco, N., Marcinek, L., Tomoriova, B., and Hladek, L. (**2015**). "Contextual plasticity, top-down, and non-auditory factors in sound localization with a distractor," Journal of the Acoustical Society of America **137**.

Kopco, N., and Shinn-Cunningham, B. G. (**2011**). "Effect of stimulus spectrum on distance perception for nearby sources," J. Acoust. Soc. Am. **130**, 1530-1541.

Larsen, E., Iyer, N., Lansing, C. R., and Feng, A. S. (**2008**). "On the minimum audible difference in direct-to-reverberant energy ratio," J. Acoust. Soc. Am. **124**, 450-461.

Lee, J., and Groh, J. M. (**2012**). "Auditory signals evolve from hybrid- to eye-centered coordinates in the primate superior colliculus," Journal of Neurophsyiology **108**, 227-242.

Liberman, A. M., and Mattingly, I. G. (**1985**). "The motor theory of speech perception revised," Elsevier Cognition **21**, 1-36.

Loksa, P., and Kopco, N. (**2016**). "Modelling of the Reference Frame of the Ventriloquism Aftereffect," Sborník z 16.rocníku konference Kognice a Umelý život (KUZ XVI) **16**, 105-110.

Loksa, P., and Kopco, N. (**2017**). "A Model of the Reference Frame of the Ventriloquism Aftereffect using a priori bias," in *Kognícia a umelý život*, edited by D. prof. Ing. Igor Farkaš, P. RNDr. Martin Takáč, P. doc. PhDr. Ján Rybár, and M. P. Gergeľ, pp. 116-122.

Loksa, P., and Kopco, N. (**2021**). "A model of the reference frame of the ventriloquism aftereffect," (www.biorxiv.org).

Maddox, R. K., Pospisil, D. A., Stecker, G. C., and Lee, A. K. C. (**2014**). "Directing Eye Gaze Enhances Auditory Spatial Cue Discrimination," Current Biology **24**, 748-752.

Maier, J. K., McAlpine, D., Klump, G. M., and Pressnitzer, D. (**2010**). "Context Effects in the Discriminability of Spatial Cues," Jaro-J Assoc Res Oto **11**, 319-328.

McAlpine, D., Jiang, D., and Palmer, A. R. (**2001**). "A neural code for low-frequency sound localization in mammals," Nat. Neurosci. **4**, 396-401.

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

Mendonca, C., Mandelli, P., and Pulkki, V. (**2016**). "Modeling the Perception of Audiovisual Distance: Bayesian Causal Inference and Other Models," Plos One **11**.

Middlebrooks, J. C., and Green, D. M. (**1991**). "Sound localization by human listeners," Annual Review of Psychology **42**, 135-159.

Michalka, S. W., Rosen, M. L., Kong, L., Shinn-Cunningham, B., and Somers, D. C. (**2016**). "Auditory spatial coding flexibly recruits anterior, but not posterior, visuotopic parietal cortex," Cerebral Cortex **26**, 1302-1308.

Mohl, J. T., M., P. J., and Groh, J. M. (**2020**). "Monkeys and humans implement causal inference to simultaneously localize auditory and visual stimuli," Journal of Neurophsyiology **124**, 715-727.

Moriya, T. J., Groh, J. H., and Meynet, G. (**2013**). "Episodic modulations in supernova radio light curves from luminous blue variable supernova progenitor models," Astron Astrophys **557**.

Mullette-Gillman, O. A., Cohen, Y. E., and Groh, J. M. (**2005**). "Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus," J. Neurophys. **94**, 2331-2352.

Mullette-Gillman, O. A., Cohen, Y. E., and Groh, J. M. (**2009**). "Motor-related signals in the intraparietal cortex encode locations in a hybrid, rather than eye-centered, reference frame," Cerebral Cortex **in press**.

Musicant, A. D., and Butler, R. A. (**1985**). "Influence of monaural spectral cues on binaural localization," J. Acoust. Soc. Am. **77**, 202-208.

Odegaard, B., Wozny, D. R., and Shams, L. (**2016**). "The effects of selective and divided attention on sensory precision and integration," Neuroscience Letters **614**, 24-28.

Oess, T., Ernst, M. O., and Neumann, H. (**2020**). "Computational principles of neural adaptation for binaural signal integration," PLOS Comput Biol **16**.

Pages, D. S., and Groh, J. M. (**2013**). "Looking at the Ventriloquist: Visual Outcome of Eye Movements Calibrates Sound Localization," Plos One **8**.

Park, H., and Kayeser, C. (**2020**). "Robust spatial ventriloquism effect and trial-by-trial aftereffect under memory interference," Sci Rep **10**.

Phillips, D. P., and Hall, S. E. (**2005**). "Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level," Hear. Res. **202**, 188-199.

Porter, K. K., and Groh, J. M. (**2006**). "The "other" transformation required for visual-auditory integration: representational format," Progress in Brain Research **155**, 313-323.

Pouget, A., Deneve, S., and Duhamel, J. R. (**2002**). "A computational perspective on the neural basis of multisensory spatial representations," Nature Reviews Neuroscience **3**, 741-747.

Radeau, M., and Bertelson, P. (**1974**). "The after-effects of ventriloquism," Q. J. Exp. Psych. **26**, 63-71.

Radeau, M., and Bertelson, P. (**1976**). "The effect of a textured visual field on modality dominance in a ventriloquism situation," Percept. Psychophys. **20**, 227-235.

Rayleigh, J. W. S. (**1907**). "On our perception of sound direction," Philosophical Magazine **XIII**, 214-232.

Razavi, B., O'Neill, W. E., and Paige, G. D. (**2007**). "Auditory Spatial Perception Dynamically Realigns with Changing Eye Position," J. Neurosci. **27**, 10249-10258.

Recanzone, G. H. (**1998a**). "Rapidly induced auditory plasticity: The ventriloquism aftereffect," Proceedings of the National Academy of Sciences **95**, 869-875.

Recanzone, G. H. (**1998b**). "Rapidly induced auditory plasticity: The ventriloquism aftereffect," P Natl Acad Sci USA **95**, 869-875.

Roseboom, W., Kawabe, T., and S., N. (**2013**). "The cross-modal double flash illusion depends on featural similarity between cross-modal inducers," Sci Rep **3**.

Salminen, N. H., May, P. J., Alku, P., and Tiitinen, H. (**2009**). "A population rate code of auditory space in the human cortex. ," Plos One **4:e7600**.

Shinn-Cunningham, B. G. (**2000**). "Adapting to remapped auditory localization cues: A decision-theory model," Percept. Psychophys. **62**, 33-47.

Shinn-Cunningham, B. G., Kopco, N., and Martin, T. J. (**2005**). "Localizing nearby sound sources in a classroom: Binaural room impulse resonses," J. Acoust. Soc. Am. **117**, 3100-3115.

Shinn-Cunningham, B. G., Santarelli, S., and Kopco, N. (**2000**). "Tori of confusion: binaural localization cues for sources within reach of a listener," J Acoust Soc Am **107**, 1627-1636.

Schiller, P. H., True, S. D., and Conway, J. L. (**1979**). "The effects of frontal eye field and superior colliculus ablations on eye movement," Science **206**, 590-592.

Stecker, G. C., Harrington, I. A., and Middlebrooks, J. C. (**2005**). "Location Coding by Opponent Neural Populations in the Auditory Cortex," PLoS Biology **3**, e78.

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

Stern, R. M., and Shear, G. D. (**1996**). "Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay," J. Acoust. Soc. Am. **100**, 2278-2288.

Taboga, M. (**2017**). "Normal distribution - Maximum Likelihood Estimation," in *Lectures on probability theory and mathematical statistics, Third edition. Kindle Direct Publishing.*

Tong, J., Li, L., Bruns, P., and Roder, B. (**2020**). "Crossmodal associations modulate multisensory spatial integration," Atten Percept Psycho **82**, 3490-3506.

Urbantschitsch, V. (**1888**). "Ueber den Einfluss einer Sinneserregung auf die übrigen Sinnesempfindungen," Pflüger, Arch. **42**, 154-182.

Van Barneveld, D., and Van Wandrooij, M. (**2013**). "The influence of static eye and head position on the ventriloquist effect," Eur J Neurosci **37**, 1501-1510.

van der Heijden, K., Rauschecker, J. P., de Gelder, B., and Formisano, E. (**2019**). "Cortical mechanisms of spatial hearing," Nature Reviews Neuroscience **20**, 609-623.

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.

Wallace, M. T., and Stein, B. E. (**1994**). "Cross-modal synthesis in midbrain depends on input from cortex," J. Neurophys. **71**, 429-432.

Watson, D. M., Akeroyd, M. A., Roach, N. W., and S., W. B. (**2019**). "Distinct mechanisms govern recalibration to audio-visual discrepancies in remote and recent history," Sci Rep **9**.

Werner-Reiss, U., Kelly, K. A., Trause, A. S., and Underhill, A. M. (**2003**). "Eye position affects activity in primary auditory cortex of primates," Current Biology **13**, 554-562.

Wightman, F. L., and Kistler, D. J. (**1997**). "Monaural sound localization revisited," J. Acoust. Soc. Am. **101**, 1050-1063.

Woods, T. M., and Recanzone, G. H. (**2004**). "Visually Induced Plasticity of Auditory Spatial Perception in Macaques," Current Biology **14**, 1559-1564.

Wozny, D. R., and Shams, L. (**2011**). "Recalibration of Auditory Space following Milliseconds of Cross-Modal Discrepancy," J. Neurosci. **31**, 4607-4612.

Yao, L., and Peck, C. K. (**1997**). "Saccadic eye movements to visual and auditory targets," Exp Brain Res **115**, 25-34.

Zatorre, R. J., Bouffard, M., Ahad, P., and Belin, P. (**2002**). "Where is 'where' in the human auditory cortex?," Nat Neurosci **5**, 905-909.

Zierul, B., Roder, B., Tempelmann, C., Bruns, P., and Noesselt, T. (**2017**). "The role of auditory cortex in the spatial ventriloquism aftereffect," Neuroimage **162**, 257-268.

Zwiers, M. P., Versnel, H., and Van Opstal, A. J. (**2004**). "Involvement of monkey inferior colliculus in spatial hearing," J Neurosci **24**, 4145-4156.