PAVOL JOZEF ŠAFÁRIK UNIVERSITY IN KOŠICE
FACULTY OF SCIENCE

LEARNING AUDITORY DISTANCE PERCEPTION: EXPERIMENTAL STUDIES AND COMPUTATIONAL MODELS

2015 **Ing. Ľuboš HLÁDEK**

PAVOL JOZEF ŠAFÁRIK UNIVERSITY IN KOŠICE
FACULTY OF SCIENCE

LEARNING AUDITORY DISTANCE PERCEPTION: EXPERIMENTAL STUDIES AND COMPUTATIONAL MODELS

DOCTORAL THESIS

Study program:	9.2.1 Computer Science
Department:	Institute of Computer Science
Supervisor:	doc. Ing. Norbert Kopčo, PhD.

Košice 2015

Ing. Ľuboš HLÁDEK

Acknowledgement

I want to thank my Savior Jesus Christ, my dearest wife Maya without whom this work would never be finished, my parents Daniela Hládeková and Jozef Hládek for their love, psychological, and material support during the years of my studies. I also want to thank my sister Katarína Hládeková and brother Daniel Hládek, and their families for being my good friends.

This project was done under supervision of Norbert Kopčo to whom I want to thank for his guidance, patience, and insightful commentaries. This project was done as a part of the international collaborative effort with the University of California, Riverside. I visited several times the laboratory of prof. Aaron Seitz, whose work and ideas provided new perspectives on my research. The first of the two reported studies was piloted and designed by Norbert Kopčo and Aaron Seitz. The graduate students from UCR and Beáta Tomoriová helped with the data collection. The portions of the project were published in the international conferences and the co-author on some of the publications was also Christophe C. Le Dantec, who was helping with the audio-visual study in its early stage.

This research was supported the funding from P. J. Safarik Univeristy (VVGS-PF-2011, VVGS-PF-2013-82), the EU Marie Curie project (FP7-247543), Scientific grant agency of the Ministry of Education and the Slovak Academy of Science (VEGA-1/0492/12), the Slovak Research and Development Agency (APVV-0452-12), the EU projects SOFOS (ITMS: 26110230088), and TECHNICOM (ITMS: 26220220182) funded by the ERDF.

Abstract

Auditory distance perception is crucial in many everyday situations. Acoustical cues to auditory distance vary from one environment to another and auditory system must adapt to the new acoustical scenes. This thesis presents the results of two behavioral experiments and a modeling study that examined two types of learning in auditory distance perception. The first study focused on spontaneous learning that occurs when the listener is exposed to a new reverberant environment over several days. It aimed to test whether auditory distance perception refines after spontaneous learning when the sound level cues are made unreliable in a distance localization task and the listener is forced to rely only on the reverberation-related cues. Auditory distance perception improved over seven days of training. The subjects learned more when the sound level systematically varied with the distance of auditory targets. A plausible explanation is that the subjects were using both room reverberation and sound level cues even when the cues were congruent. The second study examined visually guided recalibration of auditory distance perception, by examining the ventriloquism effect and aftereffect in the distance dimension. The results showed that there is an asymmetry between inducing the recalibration by using closer vs. farther visual adaptors. The asymmetry was largely related to the compression observed in localization of the audio-visual stimuli that were aligned in distance. A linear weighted model showed that the ventriloquism effect in distance can be explained as a combination of the visual component and auditory components assuming that the auditory component is weighted more than an optimum model would predict. These results provide further insight into perceptual mechanisms used by the brain to cope with new stimuli and environments in distance dimension.

Keywords: ventriloquism, room, sound level

Abstrakt v slovenskom jazyku

Vnímanie sluchovej vzdialenosti je kľúčové v mnohých každodenných situáciách. Akustické faktory, ktoré ovplyvňujú vnímanie sluchovej vzdialenosti sú rôzne v každom prostredí. Sluchový systém sa na tieto nové prostredia musí adaptovať. V tejto dizertačnej práci sa nachádzajú výsledky dvoch behaviorálnych experimentov a jednej teoretickej štúdie, ktoré skúmali učenie vnímania sluchovej vzdialenosti z dvoch perspektív. Prvá štúdia sa zamerala na spontánne učenie, ktoré prebieha vtedy, keď je človek vystavený novému reverberantnému prostrediu počas viacerých dní. Štúdia mala za cieľ otestovať či dôjde k zlepšeniu vnímania sluchovej vzdialenosti, keď hladina zvuku nie je spoľahlivým prediktorom vzdialenosti a subjekt je nútený určiť vzdialenosť iba na základe reverberantných informácií. Výsledky prvého experimentu ukázali, že vnímanie sluchovej vzdialenosti sa spontánne zlepšilo po siedmych dňoch tréningu. Ak hlasitosť zvukov sa systematicky menila so vzdialenosťou, došlo k efektívnejšiemu učeniu. Dôvodom môže byť fakt, že subjekty používali oba zdroje informácií (o hlasitosti i o reverberácií) i v tom prípade keď boli tieto informácie kongruentné. Druhá štúdia skúmala kontrolované učenie s vizuálnym komponentom pomocou „bruchovraveckého efektu“ a „bruchovraveckého afterefektu“ vo vzdialenosti. Výsledky ukázali asymetriu v audio vizuálnej rekalibrácií medzi bližšími a vzdialenejšími vizuálnymi adaptormi vzhľadom na vzdialenosti cieľového zvukového komponentu. Táto asymetria však bola do veľkej miery súvisela s kompresiou sluchovej lokalizácie audio-vizuálnych stimulov, ktoré boli zarovnané vo vzdialenosti. Lineárny vážený model ukázal, že audio-vizuálna integrácia vo vzdialenosti môže byť vysvetlená ako kombinácia vizuálneho komponentu a sluchového komponentu, ktorý má vyššiu váhu ako pri optimálnej kombinácií. Tieto výsledky spolu ponúkajú hlbší vhľad do perceptuálnych mechanizmov, ktoré používa mozog pri spracovaní nových stimulov a adaptácií na nové prostredia vo vzdialenostnej dimenzii.

Kľúčové slová: audio-vizuálna integrácia, miestnosť, hladina zvuku

Table of Contents

List of Figures

List of Tables

List of Abbreviations

F	**F**ixed level
R	**R**oved level
A	**A**uditory
V	**V**isual
m	**M**eter
cm	**C**enti**m**eter
ms	**M**illi**s**econd
AV	**A**udio-**V**isual
VE	**V**entriloquism **E**ffect
VAE	**V**entriloquism **A**fter**e**ffect
RM	**R**epeated **m**easures
DRR	**D**irect-to-**R**everberant Energy **R**atio
ITD	**I**nteraural **T**ime **D**ifference
ILD	**I**nteraural **L**evel **D**ifference
dBA	**d**eci **B**ell **A**-weighted
SOA	**S**timulus **O**nset **A**synchrony
JND	**J**ust **N**oticeable **D**ifference
RIR	**R**oom **I**mpulse **R**esponse
MLS	**M**aximum **L**ength **S**equence
MLE	**M**aximum **L**ikelihood **E**stimation
SEM	**S**tandard **E**rror of the **M**ean
HRTF	**H**ead-**R**elated-**T**ransfer-**F**unction
BRIR	**B**inaural-**R**oom-**I**mpulse-**R**esponse
fMRI	**f**unctional **M**agnetic **R**esonance **I**maging
ANOVA	**An**alysis **o**f **Va**riance

List of Terms

Ventriloquism Effect is an illusion of perceiving a spatially misaligned audio-visual stimulus as a single object.

Ventriloquism Aftereffect is a form of a rapid perceptual plasticity. The shift on the auditory map induced by the ventriloquism effect usually persists seconds to minutes after discrepant audio-visual stimulation.

Introduction

In cognitive science, great deal of work has been done in understanding perception and cognition. The research focused on the mechanics of the eye, ear, skin, and tongue. The finest details were discovered of how ion channels open and how intercellular fluid flows in and out when the neuron is firing. Investigating the smallest details of the brain is a legitimate approach but understanding the complex system requires also a 'view from above'. Marr and Poggio (1977) in the essay 'From understanding computation to understanding neural circuitry' states:

> Each level of description has its place in the eventual understanding of perceptual information processing and it is important to keep them separate. Too often in attempts to relate psychophysical problem to psychology there is confusion about the level at which a problem arises – is it related to mainly to biophysics (like after-images) or primarily to information processing (like to ambiguity of Necker cube)? More disturbingly, although the top level is the most neglected, it is also the most important. This is because the structure of the computations that underly perception depend more upon the computational problems that have to be solved than on the particular hardware in which their solutions are implemented.

The current study adopts this approach and it investigates the computational aspects of spontaneous learning and visually guided learning in auditory distance perception. In literature, the term 'learning' is used in various context. It often stands for classical conditioning, instrumental conditioning, reinforcement learning, or perceptual learning (Weinberger 2015). In this thesis it will refer to the capability of auditory system to change the perceptual representation of auditory space due to extensive training or direct stimulation with cognitively salient stimuli from different modality.

Spatial perception is a special chapter of auditory cognitive neuroscience. Within this chapter the horizontal and vertical sound localization was classically studied, while researchers paid just little attention to distance (Blauert 1997). Nevertheless, distance of the sound source is as important as horizontal and vertical localization, especially (1) when the source of the sound is visually or acoustically obscured, (the object is far behind

the view, or the distracting sounds are in front of or behind the target) (2) when the subject is visually impaired, (subjects with the complete vision loss were shown to learn to use distance information for orientation in space, similarly to the bat's echolocation) (3) when a tumor, stroke, or trauma affect the brain (4) in virtual reality systems and hearing aids (in virtual environments and hearing aids sounds are often localized inside the head; hearing aids, autonomous systems, or conference systems may need to estimate the distance of the sound source to enhance performance) (5) to understand learning mechanisms in the auditory spatial perception, with potential for novel clinical applications (6) to understand human cognition and its every aspect, which poses new questions for related or unrelated fields as well as it can also provide unexpected answers.

The document is sectioned into chapters starting with the state of the art of auditory distance perception, learning auditory spatial perception, and audio-visual integration (Sec. 1). The first study (Sec. 2) describes the experiment in which subjects underwent seven days of continuous auditory distance localization training. The second study (Sec. 3) reports the audio-visual training experiment in distance with congruent and incongruent audio-visual stimuli. The next chapter (Sec. 4) presents a model of cross-modal integration and compares the predictions to the results in the audio-visual study. Conclusions (Sec. 5), Resumé in Slovak language, (Sec. 6), and Bibliography (Sec. 7) constitute the final part of the document.

1 Background

1.1 Sound Localization and Spatial Hearing

Sound localization is an integral part of life in everyday situations. Understanding a friend in noisy cantina, or avoiding an approaching car would be difficult without the ability to localize the sound source and separate it from the distractor sounds. A difficulty of the sound localization is that the signals from various sources interact with each other as well as the signals are corrupted by acoustical properties of the environments on its way from the source to the ear drum. Thus the auditory system receives only a mixture of the distorted signals. Its task is to solve the 'ill-posed' problem and reversely extract the position of the sound sources. Much of the previous work has been focused on the direction of the sound in horizontal (azimuth) and vertical (elevation) dimensions. However, fewer studies investigated how people determine the distance of the sound source (Zahorik et al. 2005; Middlebrooks and Green 1991; Moore 2012).

The sound localization has been classically studied in an anechoic environment with a single sound source (Blauert 1997). The classical studies characterized the essential properties of the sound localization and they found that sound localization is very accurate in horizontal plane. The minimum audible angle in front of the listener is about 1° and it decreases with increasing lateral angle up to 10° at the side of the listener (Mills 1958). In vertical plane the sensitivity is lower than in the horizontal plane. However, localization of sounds in vertical planes strongly depends on the frequency content of the signal because it was observed that the sounds with the narrow-band frequency content (e.g., tones) cannot be localized in the vertical plane in anechoic conditions reliably while the localization of tones in horizontal plane is very precise (Blauert 1997). To localize the sounds in the horizontal plane, the auditory system uses the information from the signal disparity of time and level between the ears, so called interaural time difference (ITD) and interaural level difference (ILD). The ITD and ILD are the essential cues for horizontal sound localization; however, given the size of the human head, ITD are most effective for signals below 1.5 kHz and ILDs are most effective for signals above 1.5 kHz, which is often called the "duplex theory" (Rayleigh 1875; Macpherson and Middlebrooks 2002). On the other hand, in vertical plane (mid-sagittal plane) the auditory system relies on the change in the magnitude spectrum of the sound which systematically varies, due to the shape of head and pinnae, as the sound moves from directly ahead of

the subject to behind the subject (Musicant and Butler 1985). The effects of physical propagation of sound (reflection, refraction, and absorption) on sound localization cues in a given environment can be characterized by so called head-related-transfer function (HRTF), and binaural-room-impulse response (BRIR) (Shinn-Cunningham et al. 2005). These characteristics fully capture the acoustics of body (HRTF) and environment (BRIR) and therefore they can be used in sound reproduction or virtual reality systems.

Nevertheless, while the horizontal and vertical localization in anechoic space is relatively accurate, the studies of the distance perception (Mershon and King 1975; Mershon and Bowers 1979; Coleman 1962) observed that the judgments of the egocentric distance are highly impaired in the anechoic environment. The interaural disparities provide distance information only in the near proximity of the subject (e.g., ILD systematically varies as a function of distance up to 1 m) (Brungart and Rabinowitz 1999), differential attenuation of the high frequencies with respect to the low frequencies by passage through air (Coleman 1968) affects only and sounds beyond 15 m (Blauert 1997). Thus in many situations these cues are unavailable to the listener. Therefore the main cue to the distance of the sound source in anechoic space is sound level. However, it only provides relative information about the distance of the sound source. To be able to use the sound level as the cue for auditory distance, the subject must have a prior knowledge (Wisniewski et al. 2012; Coleman 1962) about how loud the sound should be, or they must perceive the change of the sound level, either by self-movement or by the movement of the sound source (Hall and Moore 2003; Ashmead et al. 1990). It means that there exists only a limited set of cues for auditory distance in anechoic space and people cannot perceive distance of the sound source correctly (Shinn-Cunningham et al. 2000). Nevertheless, the subjects are able to perceive the egocentric distance accurately in regular environments where the sound is reflected by the surrounding surfaces because the reflections provide a salient cue (Mershon and King 1975; Mershon and Bowers 1979). The cue is the direct-to-reverberant energy ratio (DRR), which expresses the amount of energy that reaches our ears vs. the amount of energy reflected from the environment. This cue varies systematically with distance in every reverberant room. It is independent from the overall intensity, and it provides an absolute cue for distance. However, it was shown that the perception of auditory distance is influenced by the immediate experience with the room reverberation. The previous studies (Mershon et al. 1989; Coleman 1962) showed that the blindfolded subjects, who did not have any prior experience with the room reverberation, refined the perception of sound distance only

after few presentations, while no such improvement was seen in the room with limited reflections (Mershon et al. 1989; Shinn-Cunningham 2000b; Wisniewski et al. 2012) and the improvement continued over several days (Shinn-Cunningham 2000b; Kopčo et al. 2004b, 2004a; Tao et al. 2013; Chan et al. 2012a). These results suggest that the auditory system learns auditory distance cues every time when it is exposed to a new room because the acoustical properties change room to room, while in the vertical and horizontal dimensions the auditory system relies on the binaural and spectral cues that do not need to be adapted so often.

Although a single sound produces strong localization cues, in the real situations we usually hear many sounds that interact with each other. Other sounds are either produced by the different sound sources that reach our ears directly or they are echoes (the signals that were reflected and refracted by the surrounding materials). To study sound localization in presence of multiple sounds, the experiments manipulated the exact number of signals that reached the listener and their time of arrival. The studies (Blauert 1997) showed that when two displaced brief signals (clicks) reach the listener at the same time, the perceived position was a sum of the positions where the signals originated (e.g., if one signal was 45° to the left and the other signal was 45° to the right of the subjects midline, the percept was in the middle). When the stimulus onset asynchrony (SOA) of the two signals (temporal offset) increased, the percept shifted toward the leading signal. However, only when the SOA reached several milliseconds (2-5 ms for clicks) the subjects started to hear the lagging sound as a separate event, the echo pops out. The phenomenon is called the 'precedence effect' (Litovsky et al. 1999; Brown et al. 2015). These studies demonstrate the mechanism that brain uses to enhance the sound localization by suppressing the later arriving echoes in reverberant rooms. A different perspective on the same phenomena can be experienced when someone is asked to localize a tone or a narrow-band of noise in a reverberant room. Although, the tone can be readily localized in an anechoic environment, in a regular environment the task is sometimes very difficult (Hartmann 1989). It shows that when a signal is accompanied by reflections coming from various directions, the auditory system receives unreliable information about the position of the sound and it cannot assign a single location to the sound. However, when the signal is increased in bandwidth or when it contains a sharp onset, the sound becomes clearly localizable even in a reverberant room (Hartmann 1989, 1983; Rakerd 1986, 1985). This demonstrates that the auditory spatial information is integrated across many frequency channels. However, the actual mechanism is much

more complex including integration over various frequency channels and interactions on the various stages of auditory processing (Dietz et al. 2011; Faller and Merimaa 2004; Braasch 2013).

Auditory spatial processing in reverberant environments poses a great computational challenge for the auditory system, the neural structures involved in the processing of complex acoustical scenes were identified along whole auditory pathway but it is not clear whether and how these phenomena also relate to distance perception, which heavily depends on reverberant cues.

1.2 Sound Localization in Distance Dimension

Auditory distance perception can be generally characterized as imprecise with high across-subject variability. The localization error are commonly as big as 20%-50% of the target distance. Usually, the judgements are compressed; the near sounds are overestimated and the far sounds are underestimated (Zahorik et al. 2005).

Mean judgments, i.e., response accuracy, in many previous experiments (Zahorik et al. 2005) were characterized by the equation $d' = kd^a$ where perceived distance d' and presented distance d were in the power relationship. Term k expressed linear compression or expansion, and term a expresed the power compression or expansion. The usual parameters fits were $k \sim 0.15 - 0.7$ and a slightly more than 1 , which explains the compression in the subject responses.

Only few studies (Zahorik et al. 2005) investigated the localization 'blur' (the sensitivity to change of distance). The studies reported that standard deviation of response ranged from 20%-60% of reference distance. Another study (Kopčo and Shinn-Cunningham 2011) was measuring correlation coefficients of perceived vs. presented distance as a function of target laterality and frequency. The study found that the correlation coefficients were influenced by the spectral content of the stimuli (higher low-frequency cut-off decreased correlation coefficient) and the lateral sounds had slightly higher correlation coefficients than medial sounds. Although the effect of target laterality on correlation coefficients is pronounced more in the anechoic space because the ILD provides information on the side, in the midline the subjects do not have strong distance cues (Brungart and Rabinowitz 1999; Shinn-Cunningham et al. 2000). Kopčo et al. (2012) were measuring discrimination of distance perception and they found that sensitivity was constant as long as the relative difference of distance was fixed (is constant for pairs of

25cm-50cm, 50cm-100cm, etc.), which means that the sensitivity in absolute values is higher for near sounds and linearly increases with distance. Therefore the standard deviation of response should be approximately constant with increasing distance on the logarithmic scale (Kopčo and Shinn-Cunningham 2011).

The natural cues for auditory distance are sound level and reverberation (Warren 1999; Zahorik et al. 2005). The sound level provides only *relative* information about sound source distance and the listener must have a prior knowledge about the sound in order to correctly judge distance. The second most important cue is reverberation which provides absolute information about distance although there are many potential *acoustic cues*, which can be used by the auditory system to acquire a sense of depth:

1. Monaural cues
    a. Sound level (Ashmead et al. 1990; Strybel and Perrott 1984; Zahorik 2002b)
    b. Direct to reverberant energy ratio (Bronkhorst and Houtgast 1999; Kopčo and Shinn-Cunningham 2011; Zahorik 2002b; Mershon and King 1975)
        i. Amplitude modulation (Zahorik and Anderson 2014; Kuwada et al. 2015; Kim et al. 2015)
        ii. Spectral variation (Larsen et al. 2008; Georganti et al. 2013)
        iii. Spectral centroid (Larsen et al. 2008)
    c. Temporal modulation (e.g., onset time of sound) (Larsen et al. 2008)
    d. Spectral content of sound (Fluitt et al. 2013; Blauert 1997; Little et al. 1992; Coleman 1968)
2. Binaural cues
    a. Interaural level difference (Brungart and Durlach 1999; Kopčo and Shinn-Cunningham 2011; Kopčo et al. 2012)
    b. Interaural cross-correlation (Larsen et al. 2008)
    c. Binaural spectral fluctuations (Georganti et al. 2013)
3. Motion cues
    a. Relative cues (Mershon and Bowers 1979)
    b. Motion parallax (Speigle and Loomis 1993)

c. Auditory looming and receding (Bach et al. 2009)

This is not an exhaustive list and it is out of scope of this thesis to characterize each of the cues in detail. Therefore the reader should be directed by the references and the most recent reviews (Zahorik et al. 2005; Ahveninen et al. 2014; Kolarik et al. 2015).

Auditory distance perception can be influenced also by non-acoustic factors, for example when the sound is accompanied by a visual stimulus. Although, other factors can play a role, e.g. the vocal effort. familiarity, prior expectations, visual modality (Zahorik et al. 2005; Carlile 2014) provide essential and precise information about potential auditory targets and plays crucial role in communication and orientation in everyday environments. Therefore studying the auditory and visual spatial perceptual interactions is a good starting point.

Last but not least, the response method is another factor that can influence distance perception in addition to the acoustic and non-acoustic factors. The previous studies used walking to the heard target (Ashmead et al. 1995; Loomis et al. 1998), verbal reports (Zahorik 2002b; Calcagno et al. 2012), triangulation (Min and Mershon 2005), and direct pointing (Brungart and Durlach 1999; Kopčo and Shinn-Cunningham 2011) to report auditory distance. In the present experiments, we adopted two methods closely related to the direct pointing. In the first method, subjects typed a letter or number corresponding to the perceived distance. A similar method in the horizontal localization experiment was shown (Kopčo et al. 2015) to be more reliable than hand pointing. In the second method, the subjects were directing a visual cue to the perceived auditory distance using a trackball. In a previous studies (Wozny and Shams 2011b; Seeber 2002) the responses were collected with a similar method and the studies proved its utility.

1.2.1 Sound Level

The primary cue for auditory distance is sound level although it provides only a relative information about sound source distance. Sound level in the free-field decreases by 6.02dB for each doubling of distance as predicted by the inverse square law (Zahorik 1996). The natural decrease of sound level provides a salient cue for distance given the sensitivity to change of sound intensity is about 0.5-1 dB expressed as just-noticeable-difference (JND) (Miller 1947) which predicts 5%-10% sensitivity to change of distance in free-field (Ashmead et al. 1990). However, in real environments reverberation distorts the signal therefore the predictions does not completely hold in the real scenes and perception of sound level in reverberation dramatically differs from the ideal free-field

conditions (when the sound is presented from the ideal point source) (Zahorik 1996). The relationship between perception of sound level and distance in reverberant environments is far more more complex because subjects tend to perceive equal loudness even when distance is changing, which is called loudness constancy phenomenon (Zahorik and Wightman 2001). There many acoustical factors (e.g., increase of reverberant energy with distance (Moore and King 1999), acoustical resonance) or non-acoustical factors (familiarity) which can relate to this phenomenon, however, the mechanism is determined completely. Despite that, sound level plays a crucial role in distance perception as was shown in the weighted linear model (Zahorik 2002b) in which the distance estimates were characterized as weighted combination of the reverberation and sound level cues.

1.2.2 Reverberation

The second most important cue for auditory distance perception is reverberation and it provides absolute auditory distance cue, even if the acoustical reverberant profile changes from room to room. Every sound produced in the reverberant environment is accompanied with the reflections coming from the surrounding surfaces due to their reflective, refractive, and absorptive properties. Since the room acoustic is a linear system, it can be characterized by the BRIR. When an arbitrary sound is then convolved with the BRIR, the resulting signal has the same properties as it would be recorded in that specific room, however, only at that specific position (Shinn-Cunningham et al. 2005). A sample recording of the impulse response from a small semi-reverberant room (**Figure 2-2**) shows the direct sound (the first peak) and recognizable reflections (following peaks). The shape of the peaks are influenced by the reflections, refractions, and absorption of the surfaces that interact with signal on its way to the microphone. However, if we recorded the same signal from various distances, we could observed that the energy of the direct portion varies with distance, whereas reverberant portion is approximately constant. The direct portion of the sound field follows law as in anechoic space (6.02dB loss per doubling distance), while reflected sound field can be characterized as a diffuse sound field independent from distance therefore the direct to reverberant energy ratio (DRR) provides an absolute cue for distance (except for a constant offset that is fixed in each room).

The first quantitative model of auditory distance perception in rooms (Bronkhorst and Houtgast 1999) computed a modified DRR from a time window and prior knowledge about the room acoustics. Although the model was successful in predicting subject

responses to a large extent it can not substantially explain the distance perception of signals with countinuous temporal structure. For that reasons the later models were trying to enhance the model by assuming the sensitivity to sound source direction which can separate direct and reverberant portions more effectively using the equalization-cancelation approach (Lu and Cooke 2010; Bronkhorst 2002).

Despite the fact that people are sensitive to the DRR changes (Zahorik 2002b; Larsen et al. 2008) it is not likely that the brain computes the DRR directly (Bronkhorst and Houtgast 1999; Zahorik 2002a; Larsen et al. 2008; Kopčo and Shinn-Cunningham 2011). It rather uses different acoustical parameters that correlates with DRR e.g., amplitude modulation (Kim et al. 2015), spectral variation (Larsen et al. 2008), and spectral centroid (Kopčo and Shinn-Cunningham 2011; Larsen et al. 2008), however, the concrete mechanism has not been revealed yet.

1.2.3 Neural Correlates to Auditory Distance Perception

The neurons in ventral premotor cortex were shown to be sensitive to change of auditory distance in the near-field of the listener (Graziano et al. 1999). Similarly Kopčo et al. (2012) investigated the neural representation of near-field sounds with varying intensity and distance on the side of the listener (sounds involved both DRR and ILD) and a brain area, planum temporale (in the non-primary auditory cortex), that was sensitive to auditory distance using fMRI. Another study (Altmann et al. 2013) showed right lateralized areas sensitive to auditory distance when sound intensity is an available cue. Another investigation (Seifritz et al. 2002) ecologically relevant sound such as looming and receding sounds showed activation outside the cortical areas in right parietal, motor, and pre-motor areas. The most recent investigations revealed the role of inferior colliculus (the sub-cortical structure) in processing amplitude modulation related to changes of distance (Kim et al. 2015; Kuwada et al. 2015), therefore it is likely that various neural structures are involved in distance perception the along the auditory pathway and the mechanisms are mediated by broader representations of external space.

1.3 Adaptation and Plasticity in Auditory Spatial Perception

The auditory system analyzes the spatial information in neural structures that systematically change the response when the spatial cue is changing (Grothe et al. 2010). The internal representation of external space, i.e., the neural response to the stimulus, is

learned from experience (Kacelnik et al. 2006). However, the environment is changing and the internal representation needs to be updated.

Auditory distance perception changes after the immediate experience with the room reverberation (Coleman 1962; Mershon et al. 1989). It adapts quickly even after few presentations of sounds in various distance when the subject enter the new room. Which suggests that people adapt to the acoustics of the particular room. The room short-term adaptation effects has been observed in speech perception (Brandewie and Zahorik 2010; Ueno et al. 2005; Kopčo et al. 2013) or distance localization of forward and backward speech (Wisniewski et al. 2014).

In the studies in horizontal and vertical planes, the examples of perceptual rapid adaptive changes were observed in experiments which tested a localization of a sound preceded by another sound (adaptor), which had duration of several seconds (Kashino and Nishida 1998; Dahmen et al. 2010; Carlile et al. 2001). This type of learning was attributed to the adaptive coding strategy in the auditory pathway (Dahmen et al. 2010), which means that the firing properties of sub-cortical neurons quickly changed in response to the distribution of stimuli in recent history. Auditory spatial learning on longer time scales, sometimes referred as plasticity, was observed in juvenile birds and mammals raised either with restricted access to naturally occurring auditory cues, or prims (Knudsen 2002; King et al. 2011). These studies showed the striking effect of age on learning properties. Surprisingly a number of recent studies found the high degree of plasticity to altered spatial cues in adult humans. For instance, it was observed that the insertion of a mold into pinnae degrades the sound localization in vertical dimension (Hofman et al. 1998). However, sound localization restores when the subject wears the mold for several weeks and similar degradation and improvement was observed in a study of horizontal localization in which the earplugs altered the binaural cues (Kumpik et al. 2010). Learning can be also demonstrated in auditory virtual reality systems because sound localization with non-individualized set of auditory cues (e.g., if the shape of the pinnae changed) leads to an increase in front-back confusions (Zahorik et al. 2006), which shows the importance of experience with one's own auditory cues. This means that a sudden change of the cues leads to different interpretations. The plasticity with altered visual cues was also observed in human studies. Subjects wearing compressing prisms (Zwiers et al. 2003) changed the representation of auditory space in both vertical and horizontal dimensions according to visual feedback after three days of wearing the prisms

in the regular environment. The localization was restored when the prisms were removed although the aftereffect was observed for restricted amount of time. It is not surprising that the conditions of the listener and the environment are permanently changing and the auditory system adapts to the new conditions; however, learning is the vital mechanism to cope with changes in the environment.

Studies in the adaptation to altered cues examined the mechanisms of change in spatial representations of the auditory system. Nevertheless, the aim of the current study was to examine whether adults refines perception of auditory cues after extensive training. Several studies (Wright and Zhang 2006) asked this question in terms of the horizontal localization cues several studies (Shinn-Cunningham 2000b; Kopčo et al. 2004b, 2004a; Tao et al. 2013; Chan et al. 2012a; Kolarik et al. 2013b; Eštočinová et al. 2015) investigated this type of learning in distance dimension.

1.3.1 Short-term adaptation

Perception of auditory distance is influenced by experience. It was observed that distance judgments in the unknown environment were inaccurate. However, they improved immediately after few exposures to sounds in various distances (Coleman 1962). In a different study (Mershon et al. 1989), the distance judgments were enhanced after five presentations of sounds in various distances in a live room, and no improvement was observed in the dead room (almost anechoic). These observations not only suggest a principal difference between the anechoic and reverberant rooms but they also indicate that the people adapt to the specific reverberation on a short time scale. The effect consistency of the reverberation was investigated using a virtual acoustics. Perception of auditory distance was degraded if the reverberation changed after each trial (Schoolmaster et al. 2004, 2003) and speech identification was improved if the preceding sentence carrier was presented with the same reverberant profile as the target word (Ueno et al. 2005; Kopčo et al. 2013; Brandewie and Zahorik 2010).

The mechanisms of the auditory distance learning has been investigated in several studies (Wisniewski et al. 2014). Auditory distance perception improved after a brief training although the improvements were more pronounced in forward speech stimuli than in backward speech stimuli. The event-related synchronies in various frequency bands were correlated with the performance, which suggested the involvement of the later cortical processing involved in auditory distance processing by visually impaired listeners, who may employ the later cortical processing normally assigned to the vision

perception (Tao et al. 2013; Chan et al. 2012a; Kolarik et al. 2013a). Another line of studies investigated the effect of reverberation on subcortical structures such as inferior colliculus, and showed that the coding of the binaural cues is affected by the reverberation in a negative way (Devore et al. 2009, 2010). Since the inferior colliculus is directly involved in amplitude modulation processing (Kim et al. 2015), which is one of the auditory distance cues, it is also possible that short term adaptation to reverberation emerge from subcortical processing.

In horizontal localization, the perceived direction of the sound is affected by the preceding adaptor of longer (Kashino and Nishida 1998; Carlile et al. 2001; Dahmen et al. 2010; Brown et al. 2012). The adaptor causes a perceptual shift in the position of the target in direction away from the adaptor sound, as well as to a modest change in the perceptual resolution of the probe (Dahmen et al. 2010). A similar paradigm was examined in our laboratory but the repulsive change in perceived sound location was elicited by the contextual presentation of the clicks (2 ms duration) from one a priori known location (Kopčo et al. 2007). The study revealed that the perceived position was repulsed due to the mere presence of the interleaved distractor-target click pairs (Kopčo et al. 2015) therefore it was affected by the distribution of the stimuli, a process similar to previously mentioned adaptation.

Another study (Dahmen et al. 2010) investigated the neural substrate of the adaptation in ferret's inferior colliculus. The study involved behavioral experiments of noise adaptation in ferrets and humans, and physiological measurements in ferrets. The study replicated the previous findings of the adaptation studies (Kashino and Nishida 1998; Carlile et al. 2001), and found that inferior colliculus neurons could adjust their firing properties based on the ILD distributions of the Gaussian noises. Following the adaptation to different ILD means, the inferior colliculus neurons responded in accordance with the behavioral predictions because the firing rate did not change as long as the relative disparity was constant, as well as the inferior colliculus neurons adapted to the change in the variance of ILDs. These findings supported the idea that the brain attempts to maintain the highest sensitivity in the region where the most of the stimuli occur. Although the various studies suggested that the subcortical areas as IC are primarily involved in this type of the adaptive plasticity, other studies showed the involvement of corticofugal connections (for discussion see King et al. 2011).

The adaptation mechanism that is related to the sound localization in rooms is called the precedence effect build-up (Clifton and Freyman 1997; Keen and Freyman 2009). In contrast to the precedence effect which suppresses the later arriving sounds, the precedence effect build-up influences the extent of the precedence effect, for example when the precedence effect with a classical pair of displaced clicks suppresses the later click up to 5 ms, the same stimuli after the build-up have suppression 10 ms or more. This type of adaptation is in operation each time we enter a new acoustical scene and this is when the auditory system gets adapted to the acoustics. Later arriving stimuli can be considered as echoes, however when the subject is repeatedly exposed to the sounds with echoes, after some time the echoes become inaudible. The neural mechanisms of the precedence effect build-up are largely unknown but it is likely that cortical areas are involved in this process (Sanders et al. 2011). It is currently unknown if either the precedence effect adaptation observed after presentation of a noise or the precedence effect build-up-like adaptation is involved in auditory distance perception. Nevertheless, these phenomena are potential candidates.

1.3.2 Long-term adaptation in adults

Auditory distance perception improved after five days of training (Shinn-Cunningham 2000b). The study measured sound localization in a reverberant room using the same procedure as (Brungart and Durlach 1999) who were measuring sound localization in the anechoic space using a point sound source that was manipulated by the experimenter such that the distribution of stimuli covered the space in near proximity of the listener. In her study (Shinn-Cunningham 2000b), she found that the subjects improved localization vertical and horizontal dimensions, albeit the improvement in distance judgments was much greater. As shown in data from anechoic space (Brungart and Durlach 1999; Shinn-Cunningham et al. 2000), localization of auditory distance was highly impaired, especially when ILD cues were not present. These observations are also consistent with the observations on the short time scales (Mershon et al. 1989). On the other hand, sound localization errors decreased in the reverberant room in the course of the whole experiment (Shinn-Cunningham 2000b). This suggests that the subjects learned reverberation over several days of training. Another study (Kopčo et al. 2004a) also observed the learning effect on auditory distance perception. They observed that when the training blocks were conducted on the same day, only a small improvement was visible. However, the auditory distance judgments improved more significantly between

the training sessions, which suggests a role of consolidation in learning of acoustical memories. Learning auditory distance was also observed under virtual acoustics (Kopčo et al. 2004b). The subjects were trained in distance localization task over several days. The experimenters observed the improvement in consistency of distance judgments but the learning was affected by the context in which the stimuli were presented. The amount of improvement was greater when the room reverberation was fixed as when the room reverberation varied on trial-by-trial basis. These results suggested that the context influences not only short-term adaptation but also long term plasticity. However, in both studies, the sound level of stimuli was roved on trial-by-trial basis, which posed a question whether the unavailability of the level cues facilitated subjects to focus on the reverberation related cues which provide the absolute distance cues (i.e., subjects were learning reverberation because they were not responding by the sound level).

A potential mechanism for learning auditory distance was suggested in studies of distance localization by listeners with the vision loss who may use occipital visual areas in perception of auditory space (Kolarik et al. 2015). The early-blind participants were trained over weeks in sound-to-distance judgment task with sensory substitution devices and showed substantial learning effect although the learning was observed in relative judgments rather than in absolute judgments (Chan et al. 2012a). The analysis of the brain activity using fMRI suggested that learning was mediated by the reduced activity in inferior parietal cortex (occipital area) and hippocampus, and increased activity in the frontal and temporal lobe, which suggest a broad network including the areas which are traditionally related to learning (hippocampus). A similar study (Tao et al. 2013) observed the activation of the occipital areas, superior frontal gyrus, precuneus, and precentral gyrus during sound localization of sound source distance after learning. Although the activation differed between groups of early-onset and late-onset visually impaired participants. The results suggested that auditory distance learning involves cross-modal networks such as occipital areas in early-onset visual loss, whereas processing involves prefrontal areas and visuospatial memory for those with late-onset visual loss.

In horizontal and vertical localization, a common approach to study auditory spatial adaptation is either a restriction or alteration of naturally available localization cues. This can be done either by plugging ears or providing the artificial cues via hearing aids or headphones. Almost all studies found immediate decrease in the localization performance after the treatment i.e., subject biased responses according to the manipulation. However,

the studies differed in the restoration periods (from hours to days), in the perceived aftereffects, and in the change of sensitivity to altered cues (Carlile 2014). In one study (Kumpik et al. 2010) the auditory cues were altered by insertion of the unilateral ear plug. After wearing the earplug for several days, the human subjects could restore nearly original localization abilities if they were systematically trained in localization task. However, the learning was prevented in a sub-group that received the training during a single day. The other determinant of learning was a spectral consistency of training stimuli. The adaptation was complete but only if the subjects were trained with the broadband stimuli that contained flattened amplitude spectrum not when they were trained with stimuli with unpredictable amplitude spectrum which varied on trial-by-trial basis. Together with the lack of ILD and ITD adaptation and a small aftereffect, the authors implied that the underlying mechanism affects the weighting of the spectral cues (subjects could remap the spectral cues associated with the earplug to external locations) not learning new cues per se (enhance sensorial processing). Adaptation to supernormal cues was shown in training under virtual acoustics with visual feedback (Shinn-Cunningham et al. 1998). However, the subjects could not adapt fully and exhibited the systematic bias in responses, which were not removed even after long training.

Another line of research investigated whether the auditory system can enhance perception after extensive training of binaural or spectral cues (for review see: Wright and Zhang 2006). Studies showed that localization errors decreased over five days of training on a localization task (Shinn-Cunningham 2000b), percent correct responses increased after five days of training on 4 kHz tone but not on 0.5 kHz tone, localization improved on broadband noise only in the blindfolded group (Abel and Paik 2004), and some studies showed no improvements after training (Wright and Zhang 2006). The results of another experiment in which the subjects were trained in a discrimination task also showed that people could increase their sensitivity in perception of ILD after extensive training, which also generalized to untrained standards, while long-term improvement was seen after ITD training (Wright and Fitzgerald 2001).

The improvements in various perceptual tasks after prolonged training are often related to the perceptual learning (Ahissar and Hochstein 2004; Seitz and Watanabe 2005; Hung and Seitz 2014) In the perceptual learning the training targets sensorial and early perceptual processing per se, rather the level of higher organization. For example the subjects can improve in contrast sensitivity or motion detection task, which is reflected

in the permanent change of the firing patterns in the neural structures that are very sensitive to the trained stimuli with the expectation that the training would transfer to also untrained conditions. The connection between perceptual learning and learning auditory distance perception can be found in the studies in which the visually impaired listeners were trained in localization tasks for several weeks (Chan et al. 2012b; Kolarik et al. 2013a; Tao et al. 2013). These subjects seems to use regions various regions of the brain traditionally studied in the perceptual learning in visual domain, therefore it is possible that auditory localization learning in general can be involved in these networks (i.e., refine after prolonged training).

1.4 Audio-visual integration

When sound and light are simultaneously presented the perceived position of the combined stimulus is shifted towards the visual stimulus. This is called ventriloquism effect (Jack and Thurlow 1973; Slutsky and Recanzone 2001). The name was adopted from a performing art of ventriloquism but in the auditory research literature it denotes an audio-visual integration paradigm. The ventriloquism effect was recently described by the model of optimal integration of the auditory and visual information that were weighted proportionally to the variance of their individual representations (Alais and Burr 2004). According to this model the position of the audio-visual stimulus will be determined mainly by the visual component; however, when the spatial representation of the sound is more salient then the representation of the combined stimulus is attracted by the sound.

Several studies (Recanzone 1998; Lewald 2002) observed that the perceptual displacement of auditory spatial representation induced by the ventriloquism effect persisted to the trials without the visual component. The phenomenon called 'ventriloquism aftereffect' was shown to operate on the scales of milliseconds (Wozny and Shams 2011b), seconds (Kopčo et al. 2009), up to minutes (Recanzone 1998) after the discrepant audio-visual presentation and the displacement usually reached 25%-80% of the ventriloquism effect (Kopčo et al. 2009; Bertelson et al. 2006; Recanzone 1998).

The neural correlates of the ventriloquism effect and aftereffect potentially involve the multimodal areas as frontal eye fields and superior colliculus, as well as various parts of the auditory pathway including sub-cortical structures, primary auditory cortex, or parietal cortex (Kopčo et al. 2009). A study using electrophysiology (Bruns et al. 2011), observed that the ventriloquism aftereffect manipulated the early cortical processing –

representation of auditory space per se. The adaptation was mediated by the error signal originating in the later cortical processing – the mechanism of the ventriloquism effect.

The ventriloquism effect in distance has been traditionally related to the 'visual capture' phenomenon, when the visual component perceptually dominates over the auditory component. In the early study conducted in an anechoic space (Gardner 1968) all sounds seemed to originate from the nearest visual 'dummy' loudspeaker, which was placed directly ahead of them. The auditory target was placed several meters behind the dummy. However, when the subjects were allowed to move they could identify the real target, which underlines the role of vision in auditory distance perception. The visual capture, or 'proximity image effect, was later confirmed in the reverberant room too (Mershon et al. 1980). However, the study suggested a possible asymmetry in the ventriloquism effect because the visual stimuli behind the auditory target were less likely unified, as the visual stimuli in front of the target. Similar result was suggest recently in an experiment of the audio-visual integration in distance using virtual environment (Zahorik 2003) because the study also noted an asymmetry of farther vs. closer sounds.

On the other hand, reinvestigation of the 'visual capture' (Zahorik 2001) showed that the visual stimulus does not dominate the auditory percept if the subjects have enough visual cues. The congruent audio-visual stimuli improved standard deviation of response of auditory component (Zahorik 2001; Anderson and Zahorik 2014), although a similar study have not confirmed the improvement of standard deviations (Calcagno et al. 2012). The aiding effect of the audio-visual congruency have been observed in the studies in horizontal plane (Driver and Spence 1998) and it was explained as the effect of spatial attention, i.e., the attended spaced receive more cognitive resources. These effects were also observed also in the distance dimension (Chan et al. 2012b). In that study the visual and auditory targets were less accurately localized when the audio and visual components were misaligned in distances in contrast to the congruent condition. However, it is not clear whether the effect of congruency varied with the direction of the induced disparity.

The ventriloquism aftereffect in distance was investigated by (Min and Mershon 2005). In the experiment the auditory targets in distance were interleaved with the audio-visual stimuli that were either aligned in distance dimension or the distance of the visual and auditory components was disconcordant. The study investigated whether the adjacency principle can revert the disconcordant presentation. In other words, the study tested the efficacy of the aftereffect of ventriloquism in distance. The results suggested

that the visually closer adaptors tend to induce stronger aftereffect than visually farther adaptors, however from the study it is not clear how the potential asymmetry relates to the ventriloquism effect and whether the aftereffect or the asymmetry varies with distance if the relative disparity is held constant in distance.

1.5 Current study

The focus of this work is spontaneous and visually guided learning in auditory distance localization task. The previous studies of auditory distance learning (Shinn-Cunningham 2000b; Kopčo et al. 2004b) observed improvements in the auditory distance localization task over several days of training in a reverberant room but not in an anechoic room. It suggested that the subjects learned reverberation of a particular room to enhance auditory distance perception. The first experiment was designed to test the following hypotheses:

(1) During the training, the subjects will learn to use the auditory distance reverberation related cues when the sound level cue is not a reliable predictor of distance. The subjects do not learn how to use the reverberation related cues if the sound level cue is available and congruent with distance because learning was not observed in anechoic space (Shinn-Cunningham 2000b; Mershon et al. 1989) and sound level is primary auditory distance cue and should dominate the percept.

(2) The knowledge of room reverberation is independent from the availability of the sound level cues and therefore it will transfer between the conditions of availability of the level cues.

The previous audio-visual integration studies in distance (Mershon et al. 1980; Zahorik 2003) suggested that the ventriloquism effect in distance is asymmetric with respect to the direction of audio-visual disparity. The studies did not involve the systematic measures of both the ventriloquism effect and aftereffect and they have not investigated the ventriloquism and its aftereffects when the disparity is fixed relative to reference distance therefore it is not known how these effect vary with distance, how these effects relate to each other, and whether they vary with the direction of the disparity. The study was designed to test the following hypotheses:

(1) Audio-visual integration in distance is influenced by the distance of the auditory target. We expect to see more effective audio-visual integration in distance when

the visual adaptor is in front of the auditory target compared to the situation when the visual adaptor is behind the auditory target.

(2) The behavioral performance can be explained by increase of the localization blur with distance. Therefore the amount of integration relates to change of the perceptual properties with distance. That could be modeled in the framework of the weighted linear model as the optimal combination of the auditory and visual sensorial inputs (Alais and Burr 2004). However, the suggested decrease of integration of misaligned audio visual presentation in depth (Mershon et al. 1980; Chan et al. 2012b) can point to a different weighting than what the optimal model predicts.

2 Learning to Judge Auditory Distance in a Room with and without the Level Cue

2.1 Abstract

Room reverberation is a crucial factor in auditory distance perception. However, the acoustical properties of the auditory scene change every time when the listener enters a new room. Previous studies (Shinn-Cunningham 2000b; Kopčo et al. 2004b) examined how training improves sound localization in reverberant rooms over five days and observed improvement in the localization of sound in the distance dimension. However, such a trend was not observed in the anechoic chamber. The current study aims to test whether the subjects learn reverberation related distance cues when the sound level (as a distance predictor) is made unreliable, thus the subjects are forced to rely only on the reverberation cues. Thirty two volunteers completed the seven-day-long training of auditory distance localization task without feedback in two conditions – the sound level was either roved (R) on trial-by-trial basis or fixed (F) and decayed naturally with distance. Although the results showed that distance perception improved after training, the hypothesis was not confirmed. Although only improvement in the R was expected, the performance was improved in both training regimens. In addition to this, the improvement of localization coherence transferred from the F training to the R testing, which was also unexpected. The results imply that in the F condition the subjects were using both reverberant and sound level cues despite the sound level provided reliable information about the distance of auditory targets. Likely, the F condition provided a form of calibration, which facilitated learning in the R condition. The results indicate a more complex relationship between the reverberation and intensity cues in distance perception.

2.2 Background

Auditory distance perception is essential in many every-day situations. Listening to a friend in a noisy cantina (Brungart and Simpson 2002), avoiding an approaching train, or reaching to a ringing cell phone would be difficult without a sense of auditory distance. For familiar sounds, sound pressure level provides main cue for the distance of the object (Warren 1999), in reverberant spaces another prominent cue is the ratio of the direct and reverberant energy (Bronkhorst and Houtgast 1999). Although reverberation provides acoustical cues for auditory distance estimation (Shinn-Cunningham et al. 2005; Ihlefeld

and Shinn-Cunningham 2011; Georganti et al. 2013; Catic et al. 2013; Brungart and Rabinowitz 1999; Kopčo and Shinn-Cunningham 2011; Moore and King 1999; Kim et al. 2015; Kopčo et al. 2012; Catic et al. 2015; Brimijoin et al. 2013), the acoustical properties vary from room to room. This variation of the acoustical profiles poses a challenge to the auditory system to extract the distance cues. On the other hand, prior listening in the room influences speech perception (Brandewie and Zahorik 2010; Ueno et al. 2005; Kopčo et al. 2013) and repeated exposure to room reflections changes the perception of echoes (Clifton and Freyman 1997; Keen and Freyman 2009). Furthermore, distance perception improves with experience – the learning effect was shown in several previous studies (Shinn-Cunningham 2000b; Kopčo et al. 2004b, 2004a; Schoolmaster et al. 2003, 2004; Chan et al. 2012a; Tao et al. 2013; Eštočinová et al. 2015; Kolarik et al. 2013a). Learning spatial hearing requires many periods of training over several days (Kumpik et al. 2010; Shinn-Cunningham et al. 1998),.The consolidation phase is necessary for the transfer of the experience to the long-term memory (Kumpik et al. 2010; Lechner et al. 1999; Kopčo et al. 2004a).

Auditory distance perception improves immediately after exposure to the room acoustics (Coleman 1962), in another experiment the improvements were observed even after five presentation of the sound in various distances in live room while no such improvement was observed dead room (Mershon et al. 1989). Perception of auditory distance is influenced by the consistency of the acoustical context (Schoolmaster et al. 2003, 2004) and the auditory distance adaptation is influenced by the familiarity with the sound source (Wisniewski et al. 2012).

The distance perception improves after several days of training in a reverberant room (Shinn-Cunningham 2000b) while no such a trend was observed in an anechoic chamber (Brungart and Durlach 1999; Shinn-Cunningham et al. 2000). Another study (Kopčo et al. 2004b; Schoolmaster et al. 2004) showed that the learning process can be disrupted on the when the subjects were exposed to inconsistent presentation of reverberant cues. In both studies (Kopčo et al. 2004b; Shinn-Cunningham 2000b) the sound pressure level was an unreliable predictor of the auditory target distance, (i.e., the level varied on the trial-by-trial basis) therefore the subjects were focusing on the reverberation cues which could explain the learning process.

The current investigates how people create the reverberation related memories. It aims to test (H1) whether the subjects learn the reverberation related distance cues when

the sound level is made unreliable predictor of auditory target distance such that the subjects are forced to use only the reverberation cues to estimate the distance of the auditory targets. On the other hand, when the sound level cues are reliable predictor of the auditory target distance, the subjects do not learn reverberation related cues. It means that we expect that the availability of level cues can prevent room learning. The study also aimed to assess whether (H2) the knowledge of the room acoustics is independent of the availability of the level cues, i.e., whether the learning transfers from one condition to another.

In the experiment, the subjects were trained over seven days in the two regimens either with the sound level roved (R) or fixed (F). The performance was assessed by correlation coefficients of presented and perceived distances.

2.3 Methods

2.3.1 Subjects

Thirty-two out of fifty-three volunteer subjects completed the experiment. All subjects were young adults from the subject pool and participated after having read the instructions and after signing the written informed consent as approved by the University of California, Riverside Human Research Review Board. All subjects were naïve to the purpose of the experiment and had no or very limited experience with the experimental room and procedures except one (author).

2.3.2 Setup

The experiment was conducted in a small semi-reverberant room with internal dimensions 2.6 m x 3.3 m (T_{60} = 408 ms; (Brown 2002)). The room was carpeted with hard walls and the ceiling was covered with tiles. The array of the loud speakers used for the presentation was covered with an acoustically transparent cloth and ranged from 51 cm to 203 cm in front of a subject seated close to the center of the shorter wall such that the subject's ears were approximately 50 cm from the nearest wall facing the array of loudspeakers (see **Figure 2-1**). The first loudspeaker in the array visually and acoustically shadowed the first real target (which was at 69 cm). External sources of noise (e.g. an amplifier, digital processor, and computer) were located in a remote control room (background noise 35dBA SPL). Small letters ordered from A-Z and followed by numbers ordered from 1-0 were attached to a thin wooden frame above the array of loud-speakers and were slightly leveled such that the subject could clearly see all of them.

Linearly spaced letters and numbers ranged from 44.5cm to 267cm (6.35cm apart) and were clearly visible as a small table lamp provided just enough light for the subjects to see; the main lights were switched off during the experiment.

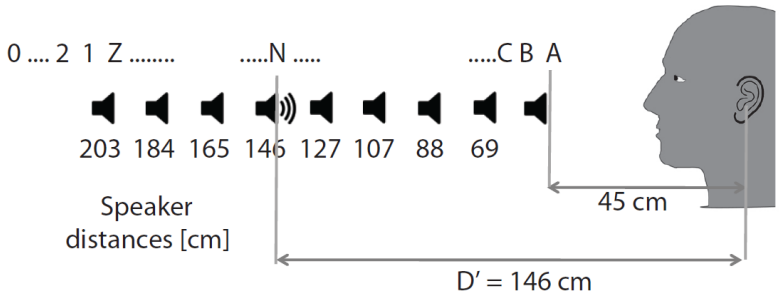

Figure 2-1 Experimental setup. Actual speaker locations and the letters/numbers (A-Z,1-0) used by listeners to indicate perceived distance. The nearest speaker was not used to present stimuli.

2.3.3 Stimuli and Procedures

The stimuli consisted of pseudo-randomly pre-generated 300 ms white-noise bursts presented from one of eight loudspeakers. White noise stimuli are often used in localization experiments in rooms because wide band spectrum provides salient localization cues (Rakerd 1986); in distance localization experiment (Kopčo et al. 2011) the consistency of responses was superior when the subjects were localizing wide-band stimuli compared to the narrow-band stimuli. In the current experiment, the subjects were instructed to indicate the position of the sound by typing the letter or the number above the array of loudspeaker. The pace of the experiment was controlled by the subject and each trial included 500 ms inter trial pause. No feedback was provided.

The noise-bursts were presented in two conditions. The first type of tokens was presented without manipulations such that the level of presentation was *fixed* (F) and the received level decreased naturally with distance. The level decreased from 56 to 53 dBA from the nearest to the farthest loudspeaker, which is less than expected from the free-field predictions and presumably relates to the short critical distance of the room (Moore and King 1999). In the second condition, the level of the tokens was *roved* (R) by 12dB (i.e., presented intensity varied from trial-to-trial) and additionally equalized according to the theoretical decay in free-field (6.02 dB per doubling distance). The experimenter instructed subjects to ignore overall level and explained the difference between conditions.

Prior to the actual experiment, hearing abilities of the participants were checked and each subject participated in a pre-training session in order to familiarize with the procedures of the experiment ("zero-day training"). Subjects started with the interval detection task, in which thresholds were estimated by the three-up-one-down staircase procedure from the last five reversals for each ear. Subjects whose threshold did not reach predefined value were excluded from the study. In the second phase, subjects were presented with stimuli used in actual experiment in the F condition with the task to report perceived distance orally with immediate feedback from a research assistant who was present in the room. In the third and the final phase, the subjects performed 320 trials of the actual task of the experiment in the F condition.

2.3.4 Acoustical Measurements

Maximum-length-sequence (MLS) is an established technique in many laboratories (Shinn-Cunningham et al. 2005) to obtain acoustical impulse response of the system and was shown to be robust to small movements and non-related noise during measuremnt (Zahorik 2002b). For the system excitation two successive 32 767 long MLS sequences were played form PC sampled at 48.828 kHz by TDT RX8 24bit A/D D/A conversion multichannel processor (Tucker Davis, Alachua, FL, USA) chained with CROW 8 channel amplifier (Crown Audio, Elkhart, IN, USA) and custom made loud-speakers mounted on the top of custom made sound absorptive stands. The same setup was used during experiments.

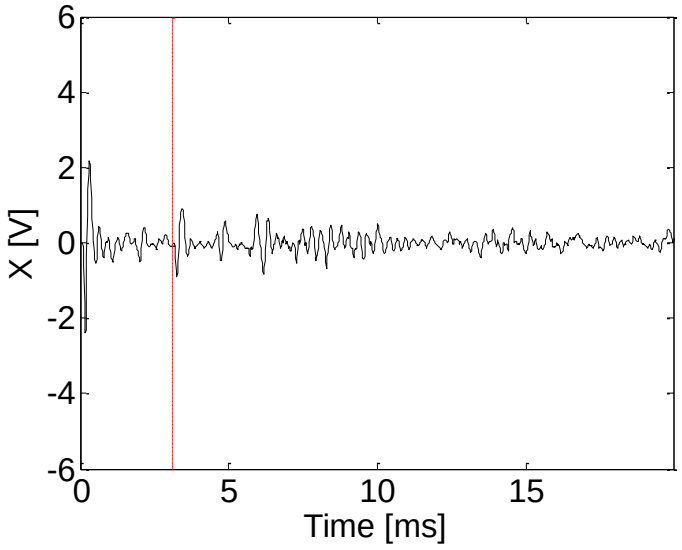

Figure 2-2 Room impulse response (RIR). An example of RIR was measured from the fourth loudspeaker at distance 1.15 m and shows the direct signals and following

reflections. The pseudo-anechoic part was defined as the first 3.1 ms of the signal when the first reflection reached the microphone.

Room impulse response (

Figure 2-2) was obtained from an average of ten measurements of MLS by inverse convolution of the original MLS. A-weighted analog output of Extech HD600 (Flir Comercial System Inc., Elkhart, IN, USA) sound meter was connected directly to the multichannel processor. To estimate the intensity of received signals, experimental stimuli measurements were performed with a ½ in. condenser free-field microphone PCB 130D20 (PCB Piezotronics, Inc., Depew, NY, USA). Microphones were placed on stand at the standard position of a subject's head, facing the array of the loudspeakers.

The microphones were placed only 0.5 m from the nearest wall therefore pseudo-anechoic part of room impulse was defined as first 3.1 ms of signal, the time when the first reflection arrived, it was shown (Shinn-Cunningham et al. 2005) that the manipulation effectively removes the effects of reverberation, although the time window in the current analysis is shorter than usual due to the room dimensions and placement of the microphone. Analysis of third-octave single-sided spectrum with respect to the mean is shown on **Figure** 2-3. The data show only small differences approximately ±3dB between speakers inside 0.2 – 10 kHz interval. The data of loudspeaker 1 are slightly misaligned but this loudspeaker was not used in the final analysis. The small discrepancies (at 230 Hz) in the data of could potentially provide some perceptual cues but the deviation was not systematically correlated with distance therefore such cues were minimized. The stimuli (300 ms white noises) were taken from the set of 50 samples.

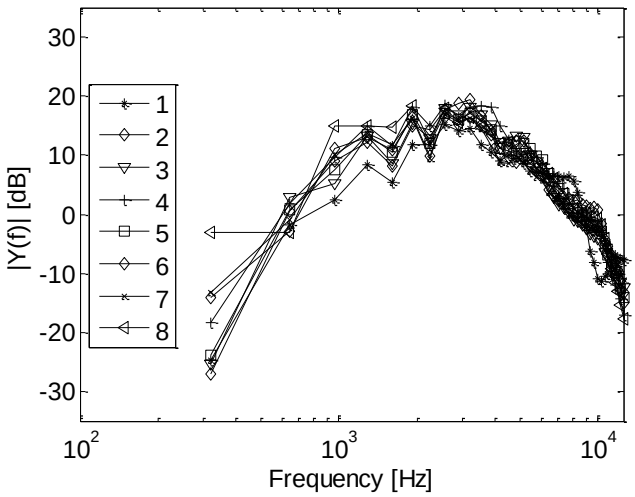

Figure 2-3 Single-sided frequency spectrum, 1/3 octave smoothed, for each target speaker estimated from the pseudo-anechoic portion of the room impulse response. Each line represents one target loudspeaker linearly spaced from 70 cm (1) to 203 cm (8).

The direct energy of the signal approximately decreased with increasing distance according to the inverse square law (-6.02 dB per doubling distance). The theoretical prediction is denoted as sloped dashed line on **Figure** 2-4. The deviations from the theoretical prediction can be attributed to the radiation of acoustic waves in the acoustical near-field and the acoustical shadowing caused by the array of the loudspeakers. Reverberant energy provided diffuse sound-filed independent of distance. The ratio of the two, difference in the current log scaled figure, defines the DRR.

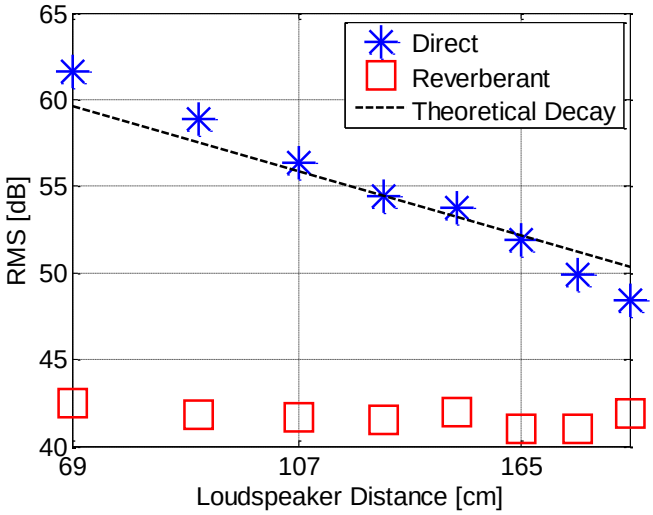

Figure 2-4 Energy the direct and reverberant portions of the room impulse response as a function of target distance. Theoretical decay in free-field is denoted by sloping black dashed line. Reverberant energy could be characterized by straight horizontal line close to zero.

Figure 2-5 shows the estimate of T_{60} (Brown 2002) for each target loudspeaker in 1/3 octave bands. T_{60} is a measure of reverberation time and expresses theoretical estimate when the energy of impulse response is decreased by 60dB. Data shown that T60 is relatively constant across target distances and frequencies at the value around 400 ms. However, there is an increasing trend at frequencies near 120 Hz in which T60 reaches almost 1000 ms. A potential explanation is that the increase relates to the room modes, a resonant room frequencies which would be consistent with dimensions of the room. The stimuli were presented along the longer 3.3 m dimension, predicts the first mode at

approximately 100 Hz (depending on the exact speed of sound), however, the increase is evident for each loudspeaker therefore it is not likely that such explanation can interfere with the reported perceptual findings. The jumps of T_{60} visible for the loudspeaker 1 and 2 most likely relate to the noise in the measurement and also are not expected to change the results.

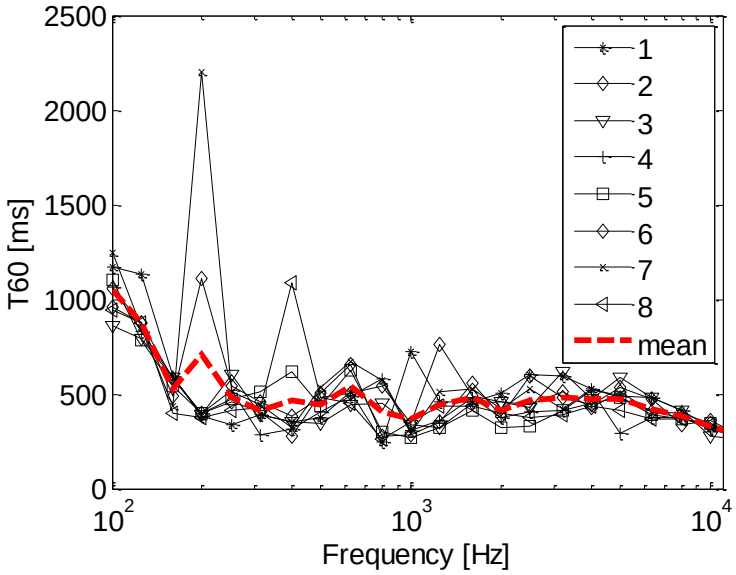

Figure 2-5 Reverberation time of the experimental room (T60) for each target loudspeaker in 1/3 octave frequency bands.

Taken together, this analysis shows that the experimental environment provided salient acoustical cues for auditory distance which were consistent with theoretical predictions. The analysis showed that DRR was present in the room but it is not likely that auditory system computes DRR from the room impulse as was shown here. Acoustical features that correlate with DRR besides intensity and are potentially used as cues for auditory distance perception in medial plane are for example spectral envelope, spectral standard deviation, temporal cues (Larsen et al. 2008), amplitude modulation (Kim et al. 2015) or binaural parameters like interaural cross-correlation (Zahorik 2009), differential spectral standard deviation (Georganti et al. 2013), interaural level difference (Brungart and Durlach 1999), interaural level and time difference fluctuations (Catic et al. 2015), or other cues. However, the experiment did not manipulate these parameters as a function of distance.

2.3.5 Experiment

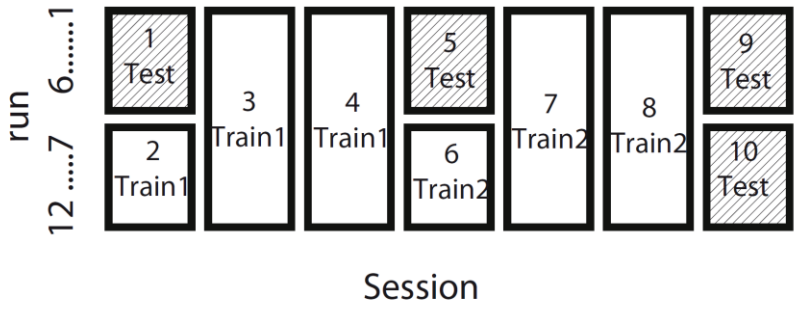

Figure 2-6 Ordering of test runs and training runs across sessions. Each rectangle represents one session, two sessions in one column indicate that the sessions were conducted on the same day. Test sessions were the sessions in which conditions altered after each run and subjects either started with R or F condition, thus Rinit or Finit groups. During training sessions the condition was fixed during whole training block i.e. (Train1=F training, Train2=R training) or (Train1=R training, Train2=F training).

The experiment was performed during 7 days in 10 sessions. Sessions were divided into 12 runs per session but if two sessions were performed on a single day (first, fourth, and seventh), sessions consisted of 6 runs. **Figure 2-6** shows the organization scheme of the experiment. *Testing* sessions 1, 5, 9, and 10 were performed at the beginning, in the middle, and at the end of the experiment and consisted of 6 runs each and the conditions *alternated* after each run. The initial condition was counterbalanced across groups forming Rinit and Finit groups of subjects. The initial order of the testing for each subject was fixed during the experiment. *Training* sessions were divided into two phases: sessions 2-4 (Train 1) and sessions 6-8 (Train 2) with the condition *fixed* within the phase. The order of the two training regimens was counterbalanced across subjects thus each subject performed either RF or FR training order, therefore each participant was randomly assigned to one of four experimental groups which differed only in the order of training sessions (FR, RF) and the initial testing condition (Rinit, Finit). Training sessions consisted of 12 runs and were conducted on separate days but sessions 2 and 6 had only 6 runs and were performed after the testing sessions.

Each run consisted of 80 trials (8 loudspeakers x 10 repetitions). Between the runs subjects were instructed to relax and the performance of the previous run was reported as

percentage expressed in Pearson's correlation coefficient of perceived vs. presented distance on logarithmic scale times 100.

2.3.6 Analysis

The localization performance was evaluated using the Spearman's correlation coefficient of perceived and presented distance computed separately for each subject and each run. Data were transformed to Z-scores using arcus tangent transformation. The Spearman's correlation coefficient formally equals to Pearson's correlation coefficient of ranked data and expresses the degree of monotonicity between two random variables. Spearman's correlation coefficient in comparison to Pearson's correlation coefficient, which expresses degree of linearity, is less sensitive to outlier measurements and can better account for high across subject variance as reported in many previous studies (Zahorik et al. 2005).

In addition to that, responses to the very first loudspeaker were omitted from the computation of the correlation coefficients in order to minimize across subject variance resulting from different response strategies observed in raw data, i.e., some subjects consistently responded to the only one location, other subjects spread their responses, which created between-subject imbalance. Altmann et al. (2013) also observed that responses to the target at a distance closer than 1 m were insensitive to intensity variation and the subjects were constantly responding to one place. This can relate either to a different response strategy, near-filed acoustics, i.e. inverse square law is violated for distances closer than 1 m, or spectral artifacts specific to experimental setup.

2.4 Results

Figure 2-7 plots performance as Spearman's correlation coefficient of perceived and presented distance of the auditory target during the whole Experiment 1 in the F condition (magenta) and the R condition (green). Testing was performed in the initial, the middle, and the final testing sessions (dark hue), always in both conditions (R,F). The testing sessions, divided the experiment in two training phases with the training condition fixed during each phase (light hue). Panels (A) – (D) show data for four subject groups (Rinit RF, Finit RF, Rinit FR, and Finit FR). The abbreviations denote the initial run of the whole experiment Rinit (dashed lines) or Finit (solid lines), and the order of training phases FR (thick lines) and RF (thin lines). Thin dashed vertical lines discriminate the

days with the testing sessions. Error bars indicate the across subject standard error of the means (SEM).

Overall, the correlation coefficients increased over the course of the training even if the learning profiles varied across the groups and conditions. In **Figure 2-7**, the majority of the improvements can be seen (1) in the first session in the first three R testing runs – (A) – (D) the first three dark green circles have an ascending trend (2) at the beginning of the R training but only when the R training was done in the first training phase – (A), (B) green light circles in session 2 are always above green dark circles in session 1; (C),(D) green light circles in the 6th session and dark green circles in session 5 are almost identical (3) between the F training sessions – jumps of the performance are visible in magenta lines always after dashed vertical lines, e.g., (A) sessions 7,8 (C) session 4 (D) sessions 3,4 (4) in the R test runs after the first training block– (A) ,(B) green dark circles in session 5 are superior to the green light circles in session 4 (5) in the F testing runs of Finit groups after the first training block – (B), (D) magenta dark circles in session 5 are always above the magenta dark circles in session 1. The error bars in Finit FR group (D) are slightly higher compared to the rest of the groups (A) - (C) but the null hypothesis that the four groups come from populations of unequal variances was rejected (Bartlet's test was performed on the averaged data of testing sessions 1,5,9; $p>0.05$).

These patterns can be summarized as improvements in the testing sessions and improvements within training sessions which will be analyzed below.

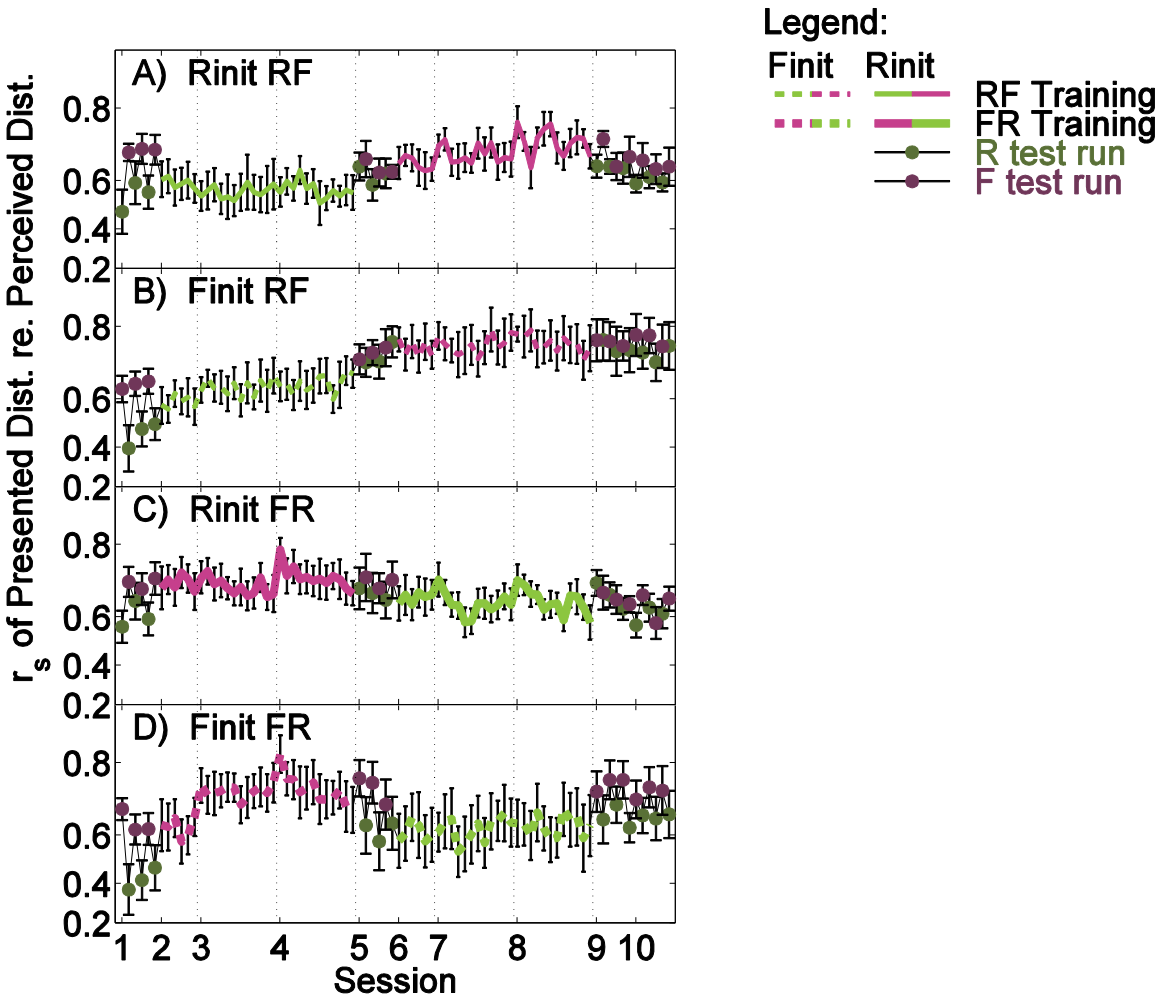

Figure 2-7 Learning curves for four subject groups in panels A) – D). Each point represents the mean value of correlation coefficient of presented and perceived distance in one run (data are taken from 70 out of 80 trials) across the whole Experiment 1. Session number is denoted on x-axis. Testing and training phases were organized as shown on Figure 2-6. During testing (dark circles), the condition of presentation alternated, whereas during the training (light circles), the presentation condition was fixed. Each subpanel includes the group name. E.g., Finit and Rinit stands for the condition at the very beginning of the experiment. FR and RF express the order of conditions in two training blocks. Vertical dashed lines differentiate the days in which the sessions were conducted. Error bars are standard error of the means (SEM) in this and all following figures.

2.4.1 Testing Sessions

To highlight the learning effects, **Figure 2-8** shows the performance only in the testing sessions. The left panel (A) shows the R testing and the right panel (B) shows the

F testing. The data are averaged across the runs and shown separately for each subject group (dark circles in **Figure 2-8** correspond to dark circles in **Figure 2-7**). The lines code the training condition and they are shown with the light color. Dots are shown with the dark color. Thus the RF training groups are coded with green-magenta compound of the line segments, while magenta-green compound stands for the FR training order. The Rinit groups are coded with the solid lines, the Finit groups are coded with dashed lines.

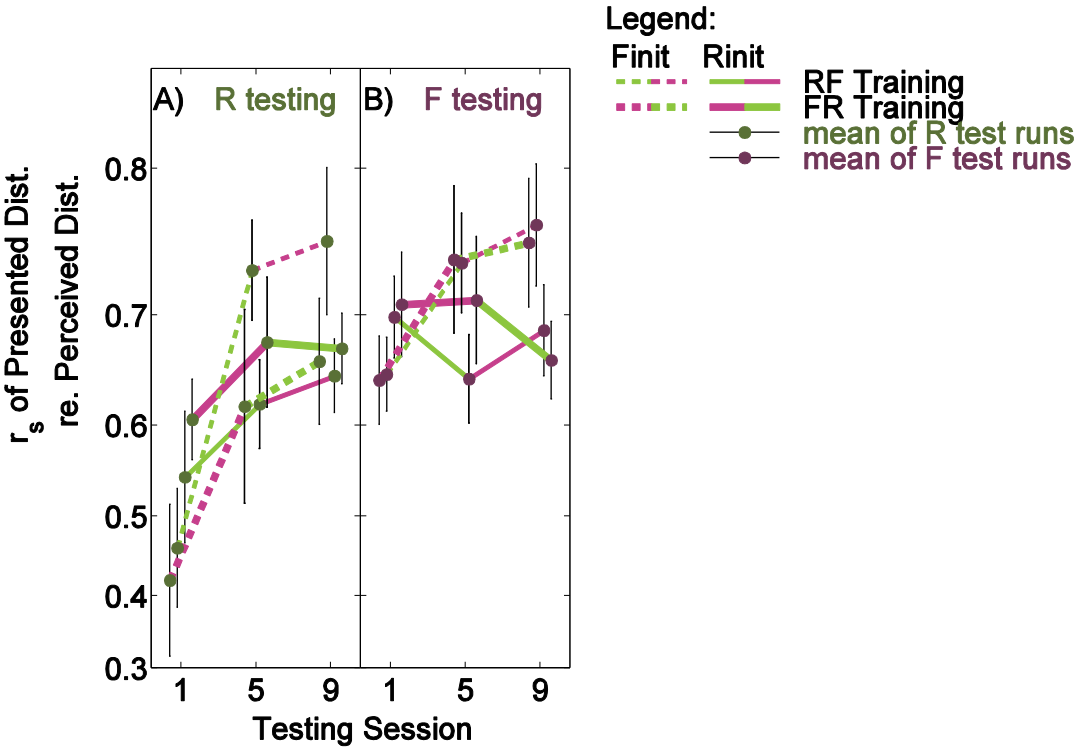

Figure 2-8 A summary of the performance in the testing sessions in the two testing conditions R testing in (A) and F testing in (B). Line color between two points denotes the type of the training (R – green, F – magenta). Solid lines represent the subjects in the Rinit group, dashed line stands for subjects in the Finit group.

(A) The R testing performance improved after both types of training even if there is a small imbalance between the effect of the R and the F training – green light lines are slightly steeper than the light magenta lines (3 out of 4 times). The improvement is evident mainly in the first training block (between sessions 1 and 5) but overall performance increases also in the second training block, except the subjects in the Rinit FR group. (B) The F testing performance improved primarily after the F training (light magenta lines are increasing). Notably, the Finit groups (dashed lines) have a tendency to improve while the Rinit groups (solid lines) had high initial performance and plateaued. **Figure 2-9** summarizes these findings and plots these observations as the total

improvement across the subject groups for the two regimens of training and two testing conditions (e.g., the left most light green bar represents the sum of improvements of the R testing after the R training in **Figure 2-9A** such that it equals sum of differences of the light green lines between sessions 1 and 5 of RF training groups and the sum of differences between sessions 5 and 9 of the FR training groups). Repeated measures ANOVA conducted on the data in **Figure 2-9** with factors of training and testing found a main effect of testing condition ($F(1,31)=15.33$, $p<0.01$) and interaction of the two factors ($F(1,31)=4.67$, $p<0.05$). The main effect of training did not reach significance ($F(1,31)=0.15$, $p>0.05$). Successively, a series of planned comparisons was performed, which showed a significant difference between R testing and F testing during the R training (t-test: $p<0.05$) but no difference between the testing conditions in the F training (t-test: $p>0.05$).

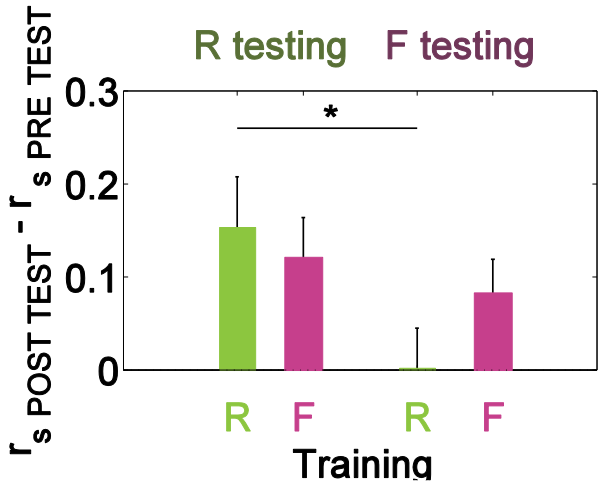

Figure 2-9 Amount of improvement due to R training vs. F training in the R testing and F testing across all subjects. The pre-test and post-test data correspond to respective test sessions in Figure 2-8, e.g. for group Finit RF, R training improvement in R testing is a difference of fifth and. first session of green dashed line in Figure 2-8A. (*two-tailed t-test: $p<0.05$)

Therefore, removing the intensity cue during the training (R training) affects the amount of learning (re. standard F training), with a positive trend in the R testing but negative trend in the F testing. On the other hand, the auditory distance training in which the sound level is correlated with distance (F training) is more beneficial than focusing only on reverberation related cues. The fact that more improvement was observed in the R testing condition than in the F testing condition can relate to the initial difference

between the R and F performance but the observed interaction cannot be fully accounted by it and suggests that the F testing benefits only little from the R training.

To further statistically evaluate the learning effects, the data from the testing sessions (dark filled circles in **Figure 2-7**) were subjected to a mixed-design ANOVA with three within-subject factors of run (1-2,3-4,5-6), session (1, 5, and 9), run type (R, F) and two between-subject factors of training order (RF, FR), and initial testing condition (Rinit, Finit). The significant main effects and interactions are summarized in **Table 2-1**. The results suggest that the most of the variance in the data can be explained by the within-subject factors and their interactions. The unexpected interaction of session and initial testing group and marginally significant interaction of training order group, initial testing group, and condition ($F(1,28)=4.19$, $p=0.050$) can explain the data only to lesser extent.

Table 2-1 Summary table of the repeated measures ANOVA on data in testing sessions with three within subject factors of testing run type (R, F), run (1-2,3-4,5-6), session (1, 5, and 9) and two between subject factors of training order group (RF, FR), and initial testing run type group (Rinit, Finit).

Factor	Df	F	
Run Type	1,28	36.18	**
Ses. x Run Type	2,56	13.20	**
Session	2,56	11.16	**
Init. G. x Session	2,56	6.62	**
Session x Run x Run Type	1,28	5.41	*

Significance levels modified by Geisser-Greenhouse epsilons: *$p < 0.05$, **$p < 0.01$.

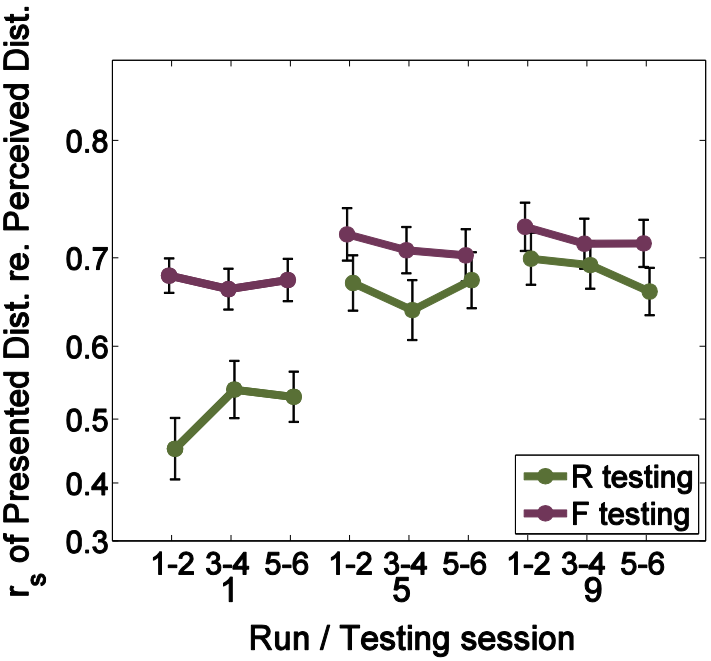

Figure 2-10 Mean testing performance. The performance is shown as a function of run and testing session for the R testing and F testing conditions. Note that R and F conditions were interleaved in the testing sessions.

Figure 2-10 visualizes the significant ($p<0.05$) three-way interaction of the factors session, condition, and run. Performance in the R testing runs are connected with the dark green lines, the F testing runs are connected with the dark magenta lines. X-axis shows six runs of the three testing sessions. The data were pooled across the training regimes and initial testing groups.

The figure shows that (1) the performance in the F test runs is superior to the R condition (2) the difference between the two conditions decreases with the training (3) there is a rapid improvement in the R testing after the first R run in the initial session. Additional ANOVA with the same design as the one above was conducted without the data of the first run. It showed both significant main effects of session ($F(2,56)= 7.94$, $p<0.01$) and run type ($F(1,28)= 25.86$, $p<0.01$), and two-way interactions of init group x session ($F(2,56)= 8.23$; $p<0.01$) and session x run type ($F(2,56)=8.12$, $p<0.1$), the three way interactions session x run x run type, and init group x run type x run were close to significance but did not reach the significance level ($p>0.05$ even without Geisser-Greenhous correction).

The difference between the R and F conditions could relate to the weighting of the level cue. Possibly the subjects were using the level cues in R condition even if it the sound level was not reliable predictor (Zahorik 2002b). The results showed that the

improvement in the R testing, regardless of the training order, was more pronounced than in the F testing. The difference may relate to the initial position on the psychometric curve, i.e. learning is faster if the initial performance is lower.

Further, data showed rapid improvement of the R testing in the first session. It is likely a form of quick calibration to the auditory space when someone enters a new room as was observed in the previous studies (Coleman 1962; Mershon et al. 1989). Although the subjects already experienced the room during the zero-day training. To investigate this particular effect of quick adaptation **Figure 2-11** shows the detail of performance in the first day of training for two groups of subjects (the training groups were identical in the first session). Besides that the figure shows the R testing (green dots) improvement after the first run, the data also suggest a slight difference between the groups (solid lines are above dashed lines). The statistical analysis RM ANOVA with factors of run (1-2,3-4,5-6), run type (R,F), and initial group (Finit, Rinit) showed the main effect of run ($F(2,60)= 2.11$, $p<0.01$), the interaction between run and run type ($F(2,60)=3.59$, $p<0.05$), and marginal main effect of group ($F(1,30)=3.84$, $p=0.0594$). The main effect of run and the interaction support the omnibus test. However, the main effect of group reveals potentially interesting finding. Since the only systematic difference between the groups was the type of the initial run, the result suggests that the calibration of distance relates to whether the sound level cues are aligned with the reverberant cues. If they are aligned, the subjects can use it as a reference of calibration, which seems to be important at the initial presentation possibly because the initial exposure is the time when the cues are actually adapted to greatest extent.

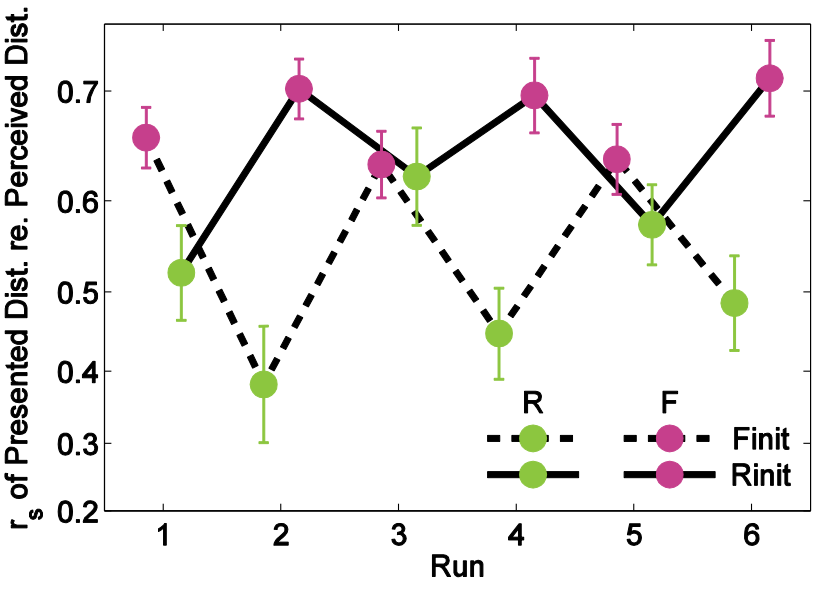

Figure 2-11 Session 1 performance of Finit (dashed line) and Rinit (solid line) groups. Data were averaged across training groups because the groups were identical in the first session.

Figure 2-12 shows the unexpected interaction of the factors of session and initial group. The mean performance is shown for the three testing sessions pooled across training order groups. Dashed line represents data of the Rinit group, solid line the Finit group. The data showed that the Rinit group started slightly below the Finit and improved during the experiment while the Finit group was superior at the beginning and plateaued in the following testing sessions. As was shown on **Figure 2-11**, the initial run of the whole experiment seems to bias the Rinit and Finit groups. The effect on **Figure 2-12** suggests that the initial bias might have persisted during the whole experiment. However, in the opposite direction as in the initial run. While the Finit subjects in the initial run improved, it seems that they also reached the maximal performance in first session. The Rinit group was initially worse and improved during the experiment. However, since this was an unexpected finding, only future investigation can reveal whether it is the 'real' effect.

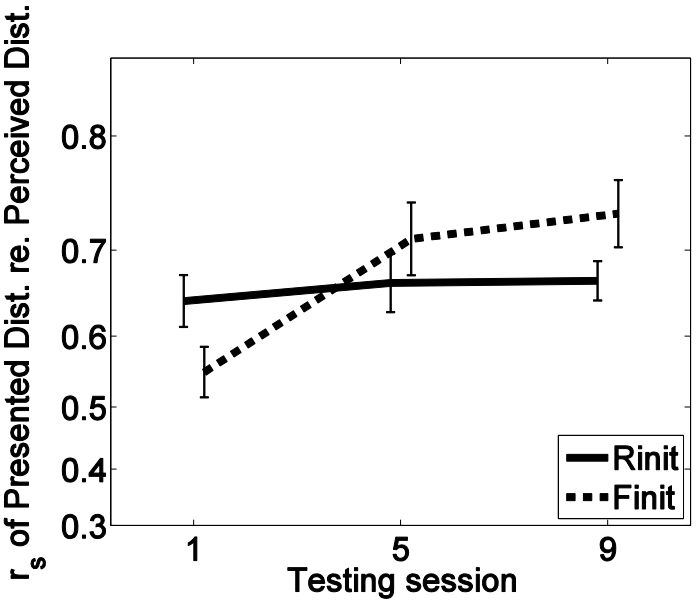

Figure 2-12 Performance in the testing sessions (1, 5, 9) according to the initial testing group (Rinit, Finit). Data are averaged across runs and training order groups.

2.4.2 Within Session and Between Session Performance

The following analysis aimed to assess the temporal profile of the two training regimens (R training and F training). Although previous analysis suggested that learning was visible mainly in the testing sessions, subjects could be learning also within training sessions.

Figure 2-12 shows the performance (A) within the training sessions and (B)-(C) performance between the training sessions. The data are pooled across training phases and subjects. (A) Solid lines show mean performance in the training sessions 3-4 and 7-8, lines with the filled circles show performance in sessions 2 and 6 as a function of run. In general, the values were constant, the magenta lines has a slight decreasing trend which means that subjects were not learning during the F training phase and the decrease can relate to fatigue. In the F condition, filled circles are slightly below the solid lines, which means that learning took place between the first and subsequent training sessions. The second panel (B) shows between-session temporal profile of the R training and F training with the pre-testing and the post-testing sessions (dark circles). Each data point shows the mean performance in one session. Panel (C) shows the temporal differences; data taken from the panel B.

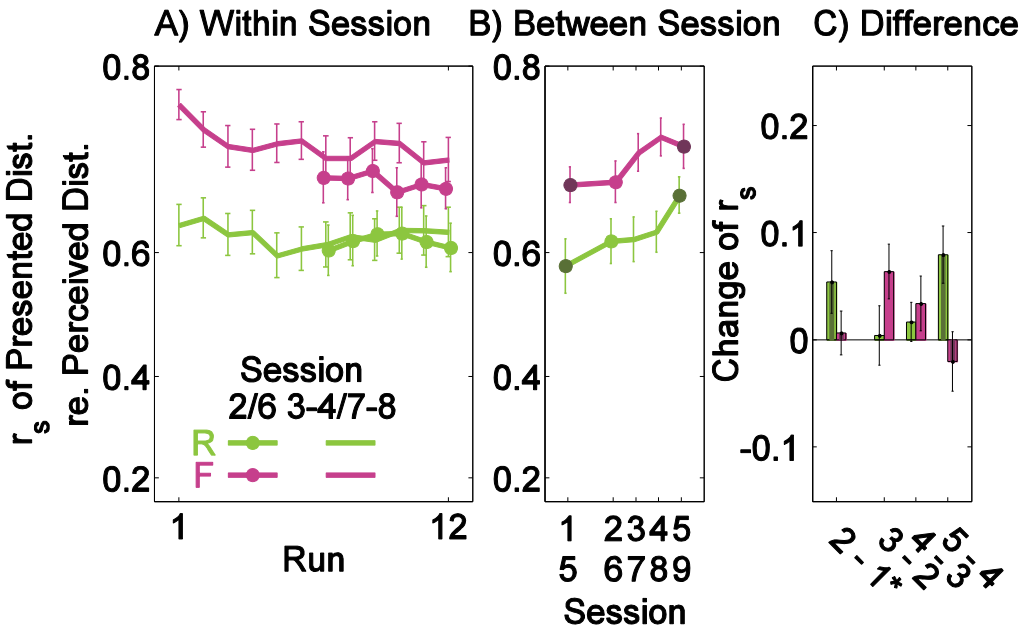

Figure 2-13 (A) Within-session performance in the training sessions (2-4,6-8) as a function of run. The training in sessions 2 and 6 had only 6 runs (light filled circles). (B) Between-session performance for two training regimens (R training - green, F training - magenta) with adjacent testing session (dark circles). The offset of the first

session means that the sessions 1 and 2 (5 and 6) were performed on a single day. (C) Change of the performance between sessions shown in middle panel (B). The data in all panels were averaged across subject groups and training phases. Caption of the x-axis was shortened for sessions 1-5 but data from sessions 5-9 were used as well as in the middle panel (B) and the right panel (C). (* sessions were conducted on a single day)

To investigate the course of the improvements during the training phase the statistical analysis assessed whether the improvements were constant during the training phase or whether the rate of improvements changed during the training phase. Therefore the improvements between the sessions of the training phase as shown on **Figure 2-12C** (sessions transitions: 3-2, 4-3, 5-4) were subjected to RM ANOVA with two within subject factors: session transition and training type (R,F). The statistical analysis showed a significant interaction ($F(2,62)=4.77$, $p<0.05$) of the two factors and no main effects. The difference in performance of the session conducted on the same day (session transition: 2-1 of **Figure 2-12C**) was assessed by a t-test ($p<0.05$).

The findings of the statistical analysis mean that there was a significant difference in learning rate between the F and R training regimens. The inspection of the graphs suggests that the F training improved mostly at the beginning of the training between sessions 2 and 3 (6 and 7) while the R training improved mostly at the end of the training between the final training session and post-testing session. These results suggest that (1) learning in the F condition happens chiefly during the consolidation between sessions (2) learning in the R condition is rapid at the beginning of training but improves during the testing. It implies that the presence of F runs is necessary for learning in the R sessions because only a small amount of learning was observed between the R training sessions without the F runs.

2.5 Discussion

The experiment showed the learning effect in the auditory distance perception after several days of training in a small reverberant room. Subjects increased consistency of responding in both training regimens, the training when the sound level provided information about distance was more effective than the training when the sound level cue was made unreliable.

The first hypothesis (H1) was not confirmed, the amount of learning in the R condition was not significantly different from the amount of learning in the F condition. The result means that at least one of three assumptions of our hypothesis does not hold.

The first assumption that subjects used reverberant cues for auditory distance perception was confirmed in many previous studies and it is not likely that the experiment proved the opposite.

The second assumption that the distance judgments are based on the level cues, when they are available, stems from the fact that the auditory system is highly sensitive to changes in sound pressure level (Ashmead et al. 1990) as well as from the fact that the sound level is considered to be the primary cue for auditory distance (Warren 1999; Zahorik et al. 2005). Furthermore, when reverberation and sound level are correlated, it is impossible to directly discriminate their influence. A study (Zahorik 2002b) investigated perceptual weighting of the sound level and reverberation cues by imposing small perturbations, which should not (in theory of weak fusion (Landy et al. 2011) disrupt the integration of the cues, i.e., the performance should be close to how people weight the cues naturally without the perturbation. The study showed that the sound level cues had significant weights although there was a substantial across-subject variation. Therefore it is likely the subjects were using both cues, sound level and reverberation, in the F condition even if they could be using only the sound level of the stimulus. The level cues provide only relative information for auditory distance. However, it seems that the subjects use both the relative and absolute information to judge distance and even further the relative information are important to calibrate the absolute distance information such as reverberation cues. Overall, this assumption does not seem to be valid, which can explain why learning was observed in the F condition.

The third assumption was that the subjects learn reverberation. This assumption was partly based on the observations that people improve performance in various perceptual tasks (Shams and Seitz 2008), including spatial perception (Wright and Fitzgerald 2001) and partly it was based on the studies that observed learning in auditory localization tasks that involved stimuli distributed in distance dimension (Shinn-Cunningham 2000b; Kopčo et al. 2004b; Schoolmaster et al. 2004, 2003). However, little is known about the actual mechanism how people store and process information about the acoustical properties of the rooms. The candidate learning mechanisms (ordered from a-e) are: (a) perceptual adaptation (Dahmen et al. 2010) (b) echo adaptation (Keen and Freyman

2009), (c) reverberation adaptation (Kopčo et al. 2013; Ueno et al. 2005; Wisniewski et al. 2014; Brandewie and Zahorik 2010) (d) perceptual learning (Ahissar and Hochstein 2004) and (e) reweighting mechanism (Kumpik et al. 2010). (a) Perceptual adaptation is a form of learning when the neural representation gets adapted to the current distribution of stimuli. This adaptation happens on the scale of seconds to minutes. (b) Echo adaptation is a quick form of perceptual adaptation that depends on the acoustical properties of the scene, and not as much on the perceptual outcome. (c) The adaptation to reverberation has been observed under various conditions (Kopčo et al. 2013; Ueno et al. 2005; Wisniewski et al. 2014; Brandewie and Zahorik 2010) and it is not strictly delimited group of observations. However, reverberation affects speech perception on the scale of seconds to minutes and it relates to the perceptual outcome, i.e., forward speech improved more than backward speech (Wisniewski et al. 2014). Although, the reported learning effects can involve multiple mechanism and that needs further investigation. (d) Perceptual learning is the adaptation of perceptual mechanism per se. It is achieved usually after extensive training (usually 5 or more days) and it affects the neural representation of sensorial processing. Perceptual learning of the room acoustical properties depends on the internal representation of the room memories, for example whether there is only one representation for all rooms, or we have specific memories for different rooms (e.g., their reverberant profiles), although neither of the two alternatives prevents from improving in one particular room. (e) Reweighting of the spatial cues is the process of changing the perceptual salience of inner representation of spatial information according to its predictive power for the current scene. In the two stage model of perceptual adaptation (Shinn-Cunningham 2000a), the spatial cues are initially extracted on the sensorial level and successively combined on the level of spatial representation. The perceptual learning would affect the sensorial processing, the reweighting mechanism would affect the spatial representation. Although it is not completely possible to assess the third assumption, the observed learning patterns can be characterized by its temporal extent. The rapid learning was observed during the first testing in R condition. The R performance improved immediately after 160 trials of practice; however the improvement was not present in the F condition. The quick improvement suggests that the learning relates either the representation of auditory space quickly adapted to the distribution of stimuli, or it relates to the reverberation learning facilitated by the presence of the F runs. With the assumption that the adaptation to a distribution of stimuli or the echo adaptation would affect both conditions because the

spatial distribution of stimuli and the room acoustics were identical in both conditions, it is more likely that the rapid improvement of R relate to the quick adaptation to reverberation. On the other hand, the slower learning was observed on the scale of days and the prominent improvements were observed between the sessions. If it were perceptual learning, the processing of reverberant information should be improved regardless of the condition, since the reverberation cues were the same in the R and F conditions. However, the R improvements were subjected to the presence of the F runs. Thus it is more likely the alternative explanation that the subjects changed the weighting of the reverberant cues (i.e., the cues that in that particular room provide more consistent information were given higher weights).

The second hypothesis (H2) was that the room learning effect did not depend on the availability of the sound level cues, i.e., the learning transfers across conditions. The results showed a significant difference between the effect of the R training and F training on the R and F testing conditions. The F training influenced both testing conditions, whereas the R training influenced only the R testing condition. Thus the result does not support the second hypothesis. The F training seems to be more effective in terms of transfer of learning than the R training. The reason can be that the room learning is facilitated by the process of calibration (the reverberation cues and the sound level are put into common reference frame), the familiarity with the range of stimulus properties, and perceptual plausibility of the stimulation (i.e., since the level cues in the F condition provided consistent information about the position of the auditory object, therefore it was more plausible for the subjects to explain the experimental presentation as one moving object,, while in the R condition the subjects could have been distracted by various potential explanations). Nevertheless, the current experiment does not favor the hypothesis that the reverberant memoires are independent from the availability of the level cues, possibly because the subjects do not learn reverberation per se rather they adapt the mapping of cues which provide consistent information in one particular room.

2.5.1 Auditory Distance Learning

The previous study (Shinn-Cunningham 2000b) found the decrease in localization error over five days of training in the R condition which is consistent with the current results. However, the present experiment also showed that the F training is equally or more effective than the R training possibly because the F condition provided a reference for the R condition. Although in the previous experiment (Shinn-Cunningham 2000b) the

sound pressure level was roved (the R condition), the perceptual reference could have been provided unintentionally by the experimenter who was present in the experimental room while the subject was responding. In our data, the R condition was improving mostly in the testing sessions in which the F runs were present. Further, the results of (Shinn-Cunningham 2000b) could be interpreted as the change of response bias. Our analysis showed that the subjects in our experiments improved consistency of responding therefore the current results provide further evidence that the internal representation is affected by learning. An important distinction between the current study and the previous study (Shinn-Cunningham 2000b) is the length of the training. The previous study trained the subjects over five days in the R condition while the subjects in the current experiment, were trained for 3 days in the R condition and another 3 days in the F condition. However, the design of the current experiment could have contributed to the difference between the experiments, especially the fact that R and F runs were interleaved during the testing sessions.

Another study (Kopčo et al. 2004b; Schoolmaster et al. 2004, 2003) was conducted under virtual acoustics in which the availability of the level and reverberation cues was well controlled. The study showed that people can improve in auditory distance judgments in the R condition. However, the study (Kopčo et al. 2004b; Schoolmaster et al. 2004, 2003) also used the process of calibration at the beginning of each run, which also supports our views. In addition to that, the experiment (Kopčo et al. 2004b; Schoolmaster et al. 2004, 2003) showed that the long-term learning can be disrupted by the inconsistent room acoustics. Similarly Kumpik et al. (2010) showed that the consistent cues prevent the long-term learning which is in line with the current recalibration hypothesis because they support the view that the context of presentation determines how the brain learns the cues of the current scene.

2.5.2 Plasticity in Vertical and Horizontal Localization

Previous studies of the plasticity in horizontal and vertical localization (Kumpik et al. 2010; Hofman et al. 1998) observed that the subjects adapted to the set of new unnatural spectral cues. The auditory system probably tries to optimize the coding of the available spatial information by adapting the firing patterns to the actual inputs (Dahmen et al. 2010) and adjust the inputs to exocentric space by reweighting of the cues. Reverberation in each room has its own acoustical profile, and if we speculate that the room acoustical profile is something similar as the new set of spectral cues when the ears

are filled with molds, or the non-individualized HRTFs in virtual acoustical space, then we can assume that people need to learn the room acoustics in each new room as it was the case in the experiment.

2.5.3 Precedence Effect Build-Up Studies

Precedence effect is a mechanism that facilitates the spatial listening in reverberant environments by suppressing the later arriving sounds (Litovsky et al. 1999; Brown et al. 2015; Keen and Freyman 2009). The studies showed that perception is calibrated after the exposure to a series of sounds compromising reflections. In connection to that, it was showed (Brandewie and Zahorik 2010) that speech perception improves if the signal is preceded by a calibration sequence. It is not fully clear whether similar mechanisms influence also auditory distance perception in the current study but Brandewie and Zahorik (2010) and other studies of precedence effect buildup (Keen and Freyman 2009) came up with the hypothesis that the model of the room acoustics is very transient, builds up and breakdowns on the scale of seconds rather than days. In contrast, the current data suggest much more complex mechanisms of room memories that can be trained over longer periods of time. However, the short term effects, as seen in the first R testing session, cannot be completely ruled out and it will need further attention.

2.5.4 Limitations

One potential limitation of the current experiment is the length of training. Our results showed only a small improvements in the R training sessions (the R improvement took place mostly between the training and testing sessions). The R performance could have been enhanced after longer training; however, that should be tested in the future studies.

Secondly, the performance in the R condition was initially worse than the performance in the F condition which may have influenced the speed and magnitude of the improvement only due to the different positions on the psychometric curve. This problem mainly interferes with the finding of the quick improvement of the R condition in the first testing session. The current design cannot separate these two alternatives sufficiently. However, the finding cannot disqualify the main findings because the F not the R condition seems to be more effective in the long-term learning, and additionally the R improvements seems to be driven by the presence of the F runs.

Thirdly, our results showed a huge variation between subjects and subject groups. The analysis tried to minimize this problem by using Spearman's correlation and excluding the first loudspeaker for the analysis. However, the presented negative results are not likely to be disqualified by a different methodology of assessing the performance. One potential source of variance was that some people were using the sound level cue as a distance predictor in the R condition. Separate analysis (Chyba! Nenašiel sa žiaden zdroj dkazov.) for people who were actually ignoring sound level for the whole experiment showed that the total change in the performance was not significant from zero (t-test: $p>0.1$), while the total amount of learning in the whole set of subjects was highly significant (t-test: $p<0.01$). Although it does not disqualify the main results because this analysis was not originally planned, this finding suggests that the perception of sound pressure level (loudness) could have interfered with our findings.

Fourthly, learning profile was unexpectedly influenced by the order of the conditions in the testing phase. Rinit group exhibited substantial learning in testing runs while the Finit group did not improve in testing runs. Most likely the F runs interfere with the ability of the subjects correctly perceive distance in the R runs, and the interaction is pronounced at the beginning of the experiment (difference between the Rinit and Finit groups). However, rather speculative reason could be that this result relates to a phenomenon known as 'ego depletion' which says that an exposure to initially demanding task can limit the assignation of cognitive resources which can interfere with the subsequent learning. For example, Thompson et al. (2014) performed an experiment in which subjects who started with a cognitively demanding task did not improve in a subsequent implicit learning task. In the current experiment, Rinit group was initially exposed to the R condition, which was more difficult than the F condition, and has not been trained in the zero-day training. Since this was a cognitively more demanding condition, subjects might have experienced learning deficits in the initial session. Although it cannot explain the long term effects as seen in the current experiment, the influence of cognitive factors on implicit acoustical learning should be examined in future studies.

3 Audio-Visual Perceptual Integration in Distance

3.1 Abstract

'Ventriloquism effect' (VE) or 'visual capture' refers to perceptual merging a sound with a visual stimulus, even when the two come from different places. Ventriloquism aftereffect (VAE) is a form of rapid plasticity of auditory spatial representation induced by the VE. This study aimed to test the efficacy of the VE and VAE with the closer and farther visual adaptors on perceived auditory distance for nearby auditory targets (in range from 70 cm - 203 cm) and when the effects were induced for a range of distances (most other studies used fixed distance). The visual disparity from the auditory target in distance dimension was either 30% closer (V-Closer), 30% farther (V-Farther), or aligned (V-Aligned). The VE was measured in the AV trials, while the VAE was measured in the A trials during adaptation while the direction of the AV disparity (V-closer or V-farther) was held constant. The VE results showed that the V-Closer stimuli were always perceived in a proximity of the visual components while the V-Farther percepts did not follow the distance of the V component and decreased significantly at distances beyond 1.5 m. However, the difference between V-Closer and V-Farther was partially explained by the compression observed in the V-Aligned condition. The VAE reached approximately 40-50% of the VE and its magnitude was constant over the range of tested distances and independent of the direction of AV disparity, suggesting that the amount of short-term adaptation is directly related to the size of the perceived, not the physical, disparity. These results provide a deeper insight into how the brain integrates information from different modalities in order to create a consistent internal representation of the world around us.

3.2 Background

When a predator hunts for prey, the estimate of the opponent's distance is critical for planning the final maneuver. Both visual and auditory information might be imprecise which leads to discrepancy in the predictions of distance, especially if only a limited number of cues are available. Moreover, the visual and auditory stimuli can be put into conflict artificially, for example in the cinema the sound is perceived as if it originated from the position of screen even if the loudspeakers are displaced from the screen. Another example is a ventriloquist's performance. The illusion of a 'talking puppet' is created by the movements of the puppet's mouth mimicking the ventriloquist's words.

The literature in the field adopted the term 'ventriloquism effect' (VE) to refer to audio-visual integration in spatial dimension (Alais and Burr 2004; Kopčo et al. 2009; Jack and Thurlow 1973; Bruns et al. 2011)

Auditory cues for distance are often imprecise and they vary between rooms. On the other hand, visual cues provide salient information. In the previous chapter (Sec. 2) it was investigated how relative cues are used for calibration of auditory distance perception. However, it is likely that subjects use also visual information as a feedback to calibrate unreliable auditory distance information.

The investigation was first conducted in an anechoic room (Gardner 1968). The subjects in that study heard a speech from loudspeaker located 9.1 m behind the 'dummy' loudspeaker such that the subjects could not see it. They reported the sound as coming from the silent loudspeaker. However, when they were allowed to move, they could perceive the sound from the correct sound source. The phenomenon was called 'proximity image effect'. Similar effect has been observed also in reverberant room (Mershon et al. 1980) because over 90% of participants reported to hear the sounds in various distances to come from the nearest visible dummy. However, the study (Mershon et al. 1980) noticed that the audio-visual unification fails more often when the visual target is farther than the auditory target, which was also confirmed by the later study Zahorik (2003) in which the sounds were presented while the listeners saw a single "dummy" target loudspeaker. The visual target 'captured' distant sounds more effectively than the closer sounds. The decrease in unification with misaligned audio-visual targets in various distances was also observed in the study (Chan et al. 2012b) in which the subjects localized visual or auditory targets, in comparison to the condition in which the the audio-visual stimuli were aligned, even if the study did not report the asymmetry in unification with respect to the direction of the induced shift. However, these studies used static visual cues and did not measure how the perception changes when the audio-visual disparity is held fixed relative to the reference distance.

The effect of visual cues on auditory distance perception in reverberant rooms was measured also in the experiments in which the availability of visual cues was manipulated (Zahorik 2001; Calcagno et al. 2012; Anderson and Zahorik 2014). In the first study (Zahorik 2001) the visual cues were restricted by blindfolding half of the participates. The results of the study showed that sighted subjects perceived distance with lower bias and standard deviation as the blindfolded group. The second study (Calcagno et al.

2012)measured auditory distance perception in a room without lights, or the light was provided by the set of LED in various distances. They observed a difference between the response biases however no difference between response standard deviations. In the third study (Anderson and Zahorik 2014), the virtual sounds were paired with the visual targets at the same distance presented on the monitor screen or the sounds and visual stimuli were presented alone. The results showed the improvement in the response consistency (measured in correlation coefficients) in auditory distance localization task when the visual cues were present with respect to auditory-only condition. Although the values did not exceed the the condition with visual-only stimuli. All these studies support the idea that visual cues calibrate auditory distance perception. Nevertheless, the audio-visual pairing and the alignment of the auditory and visual components seems to be important factors in this process.

The repeated pairings of the misaligned audio-visual stimuli produce a shift in perception of auditory target that persists from seconds (Wozny and Shams 2011b; Kopčo et al. 2009) to minutes (Woods and Recanzone 2004) after the discrepant presentation. It is called the 'ventriloquism aftereffect' (VAE). A study of VAE in distance dimension (Min and Mershon 2005) indicated that the visual adaptor placed in front of the auditory target tended to induce higher aftereffect than the visually farther adaptor in terms of bias. The study used only one adaptor for each direction of disparity and it did not measure the VE. Thus it is not known (1) what the relationship of the VAE is with respect to the VE (2) whether the magnitude of the VAE is stronger for the visual adaptors in front the auditory targets than for the visual adaptors behind the auditory targets and (3) how the aftereffect is influenced by the reference distance.

In the current study two experiments were conducted. Experiments 1 and Experiment 2 investigated immediate and persistent effects of the audio-visual training on the response bias and response standard deviation in the localization task such that the relative audio-visual disparity was held fixed. The hypotheses for the current investigation are: (1) the VE operates in distance and its magnitude is higher when the visual adaptor is placed in front of the auditory target compared to when the visual adaptor is behind the auditory target (Mershon et al. 1980; Zahorik 2003) (2) the VAE operates in distance and the VAE persists over several minutes (Recanzone 1998), its magnitude is higher when the visual adaptor (in the VE) is placed in front of the auditory target compared to when the visual adaptor (in the VE) is behind the auditory target (Min and

Mershon 2005). The aim of the experiment is also to observe the magnitudes of the effects and determine the relative magnitude of the VAE with respect to the VE. In addition to the main experiments, two supplementary experiments were conducted which measured the distance perception in unimodal conditions. Experiment 3 measured the auditory-only condition and Experiment 4 measured the visual-only condition.

3.3 General Methods

3.3.1 Subjects

Thirty four subjects participated in Experiment 1 and eighty in Experiment 2. All participants had normal or corrected-to-normal vision and normal hearing by self-report. The subjects were recruited from the university subject pool and participated only after signing the written informed consent as approved by the University of California, Riverside Human Research Review Board. All the subjects were naïve to the purpose of the experiment and had no prior experience with this or similar procedures except one.

3.3.2 Setup and Stimuli

Experiment 1 and Experiment 2 were conducted in the same room and with the similar setup as the experiment in the previous chapter (Sec. 2). Two experiments were performed in the acoustically treated room (T_{60} = 408 ms; (Brown 2002), background noise 35 dB SPL) with internal dimensions 2.6 m x 3.3 m and similar setup as the one described in the previous chapter (Sec. 2). Subjects were seated on a barber's chair with a headrest close to the center of the nearer wall facing an array of 8 uniformly spaced loudspeakers positioned in the subject's midline mounted on a custom-made stands made of sound absorbing material. The target loudspeakers ranged from 70 cm to 203 cm from subject's ears, at the height of subjects head. The array was covered by acoustically transparent cloth to minimize the visual experience with the real positions of the auditory targets. The first loudspeaker was acoustically and visually shadowed by the 'dummy' loudspeaker which was positioned approximately 50 cm from the subject ears. A wooden frame was mounted above the array of loudspeakers. 48 linearly spaced LEDs were positioned on the new frame. The LEDs were ranging from 45 cm to 272 cm of the subject egocentric distance (Figure 3-1). They were used for the presentation of the visual stimuli and colleting responses. The frame was positioned approximately 7 cm above the loudspeaker array and it was slightly slanted such that the subjects could clearly see each LED. The control PC was located in a remote room.

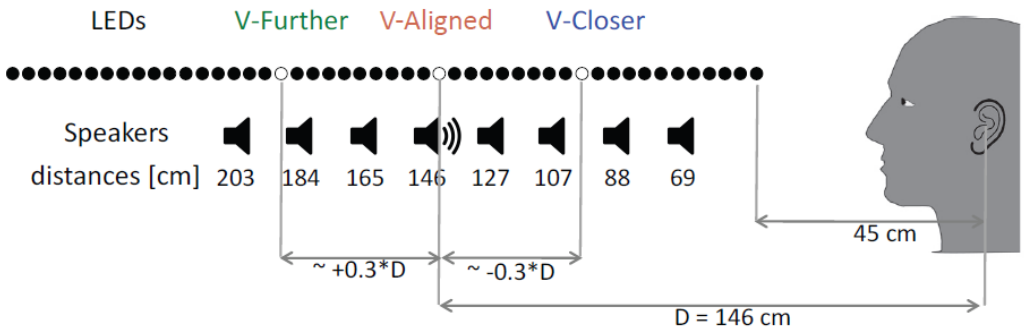

Figure 3-1 Setup of Experiment 1 and Experiment 2. Loudspeakers played 300ms white noise bursts at 53-56 dB SPL (measured at listener's position). Circles represent LEDs (open = LED on, filled = LED off). In the AV presentations, only one LED and one speaker was on at any given time. The LED was aligned with the speaker in the V-Aligned condition. In the V-Closer and V-Farther conditions, the LED was approximately 30% closer or further, respectively, than the active speaker.

The experiment involved the auditory stimuli (A) consisting of 300 ms randomly pre-generated broad-band noise bursts (53-56 dBA SPL) presented from 1 of 8 loudspeakers placed in various distances in the subjects' midline and audio-visual stimuli (AV) consisting of the A component paired with a 300 ms flash of LED light from 1 of 48 LEDs placed above the array of loudspeakers. The relative distance of the visual component was aligned (V-Aligned), 30% closer (V-Closer), or 30% farther (V-Farter) with respect to the distance of the A component. However, small deviations from the constant 30% shift occurred because of the linear spacing of the LEDs.

3.3.3 Procedures

The experiments consisted of two 1-hour-long sessions of the AV training. **Figure 3-2** shows the structure of the experiments. In Experiment 1, all subjects underwent the V-Farther and V-Closer training (V-Misaligned), in Experiment 2 each subject was randomly assigned into one of two groups. The first group underwent the V-Farther and V-Aligned training, the second group underwent the V-Closer and V-Aligned training. The order of the sessions was counterbalanced across the subjects.

Each session was executed on a separate day and consisted of 11 runs separated by short 30 seconds breaks. In the runs 1, 4-8, and 11 the AV stimuli were randomly interleaved with the A stimuli with the ratio 3:1 (75% AV, 25% A). The AV stimuli were presented in one of three conditions (V-Aligned, V-Farther, and V-Closer) with the AV disparity *fixed* during the session. The AV disparity during the adaptation (runs 4-8) was

different in each session (**Figure 3-2**). The initial (run 1) and the final (run 11) performance was assessed with zero AV disparity (V-Aligned). The pre-adaptation (runs 2-3), and the post-adaptation (runs 9-10) performance was assessed with the A presentation (A-only).

Each run consisted of 64 trials (8 target loudspeakers in pseudo-random order x 8 repetitions). After the presentation of the A or AV stimulus the subjects' task was to indicate the location where he or she heard the sound while ignoring the V stimulus. The response was collected 300 ms after the presentation of the stimulus. Random LED turned on (and stayed on) and the subject had to adjust distance of the light using a track-ball (Wozny and Shams 2011b) and submit the response using a click button on the track-ball.. During the collection phase only one LED was turned on at a time. The response was followed by 500 ms inter-trial pause.

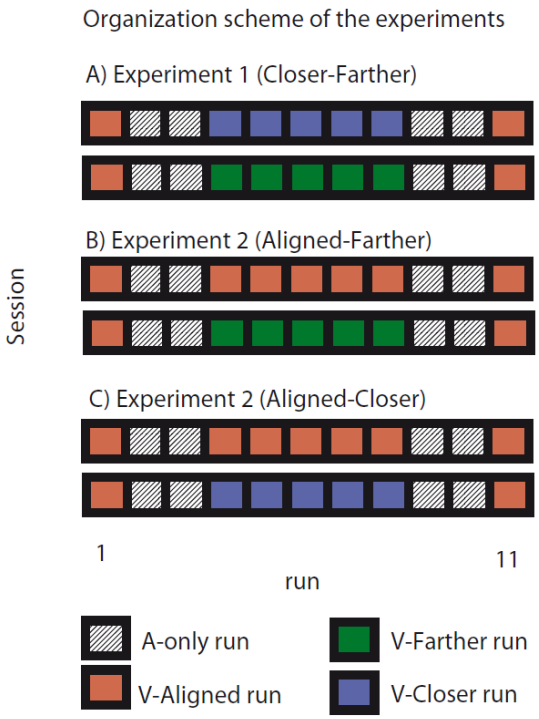

Figure 3-2 Organization scheme of Experiment 1 and Experiment 2. Two rows in each panel represent two sessions. Each block within the row represents one run. The color and hatching represent the condition. The rows show the order of experimental conditions for the experiment.

3.3.4 Analysis

The performance was assessed in terms of response bias defined as logarithm of distance minus logarithm of true distance. The within-subject response standard deviation (SD) was computed in the logarithmic units too (Kopčo et al. 2012). To compute the within-subject SD, the responses for each target were pooled across runs of identical conditions (pre-adaptation, adaptation, and post-adaptation runs). The values show across-subject mean and standard error of the mean (SEM) of these statistics.

The magnitudes of the VE and VAE were obtained by referencing the performance in the V-Misaligned by the performance in the V-Aligned condition. The VE was computed from the AV trials, the VAE was computed from the A trials. In Experiment 1, the reference was measured as across-session mean performance in run 11. In Experiment 2, each subject performed a sessions with the V-Aligned adaptation, which was used as the reference.

A simple model of the VE was considered, assuming that the VE could be explained as the portion of the complete VE (100% VE) such that the 0% VE is equal to the V-Aligned performance (i.e., light and sound are aligned at the distance of the auditory target of the VE), and 100% is equal to the theoretical V-Aligned performance at distance of the visual adaptors used in the V-Misaligned conditions. This performance is not directly measured in the experiment. It was obtained from the power model fits (Zahorik et al. 2005) of the V-Aligned performance in adaptation (in Experiment 1 runs 11) for individually each subject.

Analysis of variance with repeated measures (RM ANOVA) was conducted on biases and standard deviation. The statistical test without the description in the text involved factors of target distance (8 distances) and condition (V-Closer, V-Farther). using software CLEAVE (Herron 2005). In Experiment 1, the factor of condition was also a within-subject factor. In Experiment 2, the factor of condition was between-subject factor. The factor of target distance was always within subject factor. The reported p values of the F statistics were corrected for violations of sphericity using Geisser-Greenhouse epsilon. All other computations were done using MATLAB (MATLAB 2014a, Natick, MA, USA).

3.3.5 Experiment 3 – Auditory-only

Additional thirty-two subjects were recruited to participate in the control experiment that assessed auditory distance localization performance without the visual component.

The methods and procedures of the study were identical to Experiment 1 with the difference that subjects were localizing sounds without the visual components. The procedures followed all the ethical as the main experiments. Subjects underwent the two sessions of 11 runs (each of 64 trials) in which the subjects were localizing 300 ms broad-band noise presented in isolation, the same stimulus as was used in the main experiments.

3.3.6 Experiment 4 – Visual-only

Additional 69 subjects were recruited to participate in a control experiment that assessed visual distance perception without the auditory component. The procedures and methods were similar to the procedures and methods in the main experiments including all ethical standard. The measurement lasted only 10 minutes during which the subjects followed the identical procedure as in the A-only run of Experiment 1 and Experiment 2 with the exception that the sounds were replaced with the flashes of LEDs with identical duration (luminance was not controlled nor measured but it was the same as in the audio-visual experiments). The task was to report the perceived location of the LED flash (instead of sound). Subjects performed 80 trials such that each LED (n=48) was presented at least once.

3.4 Results: Experiment 1

Subjects in Experiment 1 were trained in the V-Closer and V-Farther conditions. The condition and relative AV disparity was held constant during the sessions. The adaptation part (runs 4-8) was the main part in which the discrepant AV training was tested. The adaptation runs were preceded by the pre-adaptation runs (2-3) and followed by the post-adaptation runs (9-10) with the A-only presentation to assess the influence of the AV training on the representation of the auditory space. To align the performance of subjects the initial run (1) was presented with the V-Aligned condition and to assess whether the aftereffect persisted to the AV run and whether the subjects returned to the initial performance, the final run (11) was also presented in the V-Aligned condition. In such way, the design preserved the temporal symmetry and the contrast effects on various timescales could have been obtained.

3.4.1 Response Bias

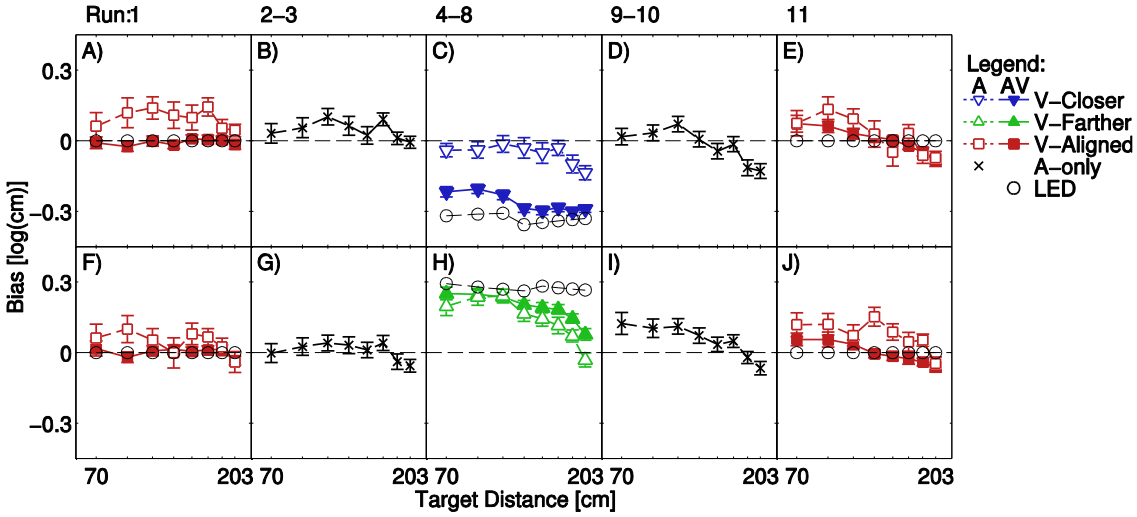

Figure 3-3 Localization bias as a function of target distance in Experiment 1. The responses in the AV trials are plotted using solid lines and filled symbols, responses in the A trials are plotted using dashed lines and open symbols. The rows represent sessions, panels C and H show the audio-visual adaptation with discrepant stimuli (V-Closer – downward-pointing-triangles), or father (V-Farther – upward-pointing-triangles) by approximately 30%. Performance in the pre-adaptation (B,G) and post-adaptation (D,I) was without the visual component and shown with the 'x' symbol. The first (A,F) and the final runs (E, J) were presented with the AV stimuli that were aligned in distance (V-Aligned). The numbers above the graph show the run numbers that were averaged in the column. The data are shown in the log-log space. Error bars show standard error of the mean (SEM).

Figure 3-3 shows the across-subject mean (± SEM) localization bias as a function of target distance during all parts of the experimental session. The middle columns (C, H) shows localization bias in the adaptation conditions with discrepant presentation which could either 30% farther (V-Farther – upward-pointing-triangles) or 30% closer (V-Closer – downward-pointing-triangles). The columns on the side of the adaptation part are pre-adaptation and post-adaptation runs with the A-only presentation ('x symbols'). The left-most and right-most column show data of the initial and the final runs with the V-Aligned (squares) condition. Two rows of the panel show the data of two sessions with different AV adaptation conditions. Solid lines with filled symbols represent mean

perceived bias in the AV trials, dashed lines with open symbols represent bias in the A trials.

Overall, the localization performance in the experiment was accurate the near targets are overestimated and the responses to far targets have tendency to be underestimated although the A trials are above the AV trials in most cases. On the other hand, the responses in the AV trials were strongly affected by the V component, the responses in the A trials were also influenced by the V component although the influence was lower.

The localization in adaptation runs (4-8) shows performance in the V-Misaligned conditions. Distance judgments were shifted in the expected directions. However, there is a clear discrepancy between the A and AV trials in the two conditions. The V-Farther A and AV trials were perceived with similar bias. The V-Closer A trials were localized almost with zero bias while the V-Closer AV trials were substantially biased in the expected direction. The performance in the pre-adaptation and post-adaptation run shows that the AV training influenced the perceptual bias in the A-only runs in the expected direction. The initial and final performance in the V-Aligned condition showed subjects could be calibrated to actual distance in the AV presentation. However, that does not transfer to the A trials. The AV trials were perceived almost with zero bias and compression in run 1, while the judgements in the A trials systematically overshot true distance. The same trend is visible in run 11. However, responses in run 11 were more compressed.

Therefore these results demonstrate that the perceptual shifts could not have been caused only by attentional factors, i.e., by the perceptual change when the visual component is present and they also provide an evidence that subjects could perform the task in the expected way.

Figure 3-4A replots the data of **Figure 3-3CH** from adaptation runs (4-8) combined into one panel. In the AV V-Farther trials (green solid line with upward-pointing-triangles), the auditory targets were perceived close to the position of the visual component at distance up to 1 m. As the distance increased, the magnitude of the AV bias decreased. In the V-Closer condition (blue solid line with downward-pointing-triangles), the bias at the nearest target distances (around 1 m) did not completely reach the distance of the visual components. The magnitude of bias increased with increasing distance. In the absolute values, the V-Closer biases were shifted more than the V-Farther biases ($F(1,33)=9.31$, $p<0.01$), the magnitude of the bias varied with the target distance

($F(7,231)=5.13$, $p<0.01$), and the interaction of the two factors reached the significance ($F(7,231)=22.79$, $p<0.01$).

In the A trials, the V-Farther responses (dashed green line with open upward-pointing triangles) were almost aligned with the AV data, while the V-Closer (dashed blue line with downward-pointing-triangles) responses were biased only by a small amount from the true target distance. Overall, the magnitude of the bias varied with the target distance ($F(7,231)=4.57$;$p<0.01$) and condition (V-Closer, V-Farther) ($F(1,33)=4.26$, $p<0.05$). The statistical analysis also showed the interaction of target distance x condition ($F(7,231)=9.19$;$p<0.01$). The V-Aligned data were obtained as an average of the performance in the final runs (11). Unexpectedly, the responses in the AV V-Aligned runs were considerably compressed (further analyzed below) even if the visual components were aligned with the auditory targets. The V-Aligned data were used as a reference for the VE and VAE.

Taken together, these results support the asymmetry of audio-visual integration in distance between the V-Closer (sound is paired with closer visual adaptor) and V-Farther (sound is paired with farther visual adaptor) conditions that was suggested in the previous experiments (Mershon et al. 1980; Zahorik 2003) because the AV V-Closer responses are very close to actual distance of visual component, while the AV V-Farther responses decrease the bias with increasing distance. It suggest that the localization was affected by the relative change of sensitivity (localization blur) of the A and V components with increasing distance. That has the assumption that the sensitivity of both components (A, V) is approximately constant on the logarithmic scale. Therefore, since the V component was farther its sensitivity decreased more (according to prediction of the logarithmic scale) than the sensitivity of the corresponding A component. That is in line with the prediction of the AV integration model (Alais and Burr 2004).

The responses in the A trials; however, do not support our predictions based on the data of the previous experiment (Min and Mershon 2005). We expected to see that A responses would be affected more by the V-Closer adaptors than by the V-Farther adaptors. However, our data suggest the opposite, that the A V-Farther responses are more influenced by the V adaptor than the V-Closer adaptor, in terms response bias with respect to true location.

However, the V-Aligned data on the figure suggest, that the observed patterns of responses, i.e., the compression and the AV vs. A difference may explain some of the variance observed in the V-Misaligned conditions.

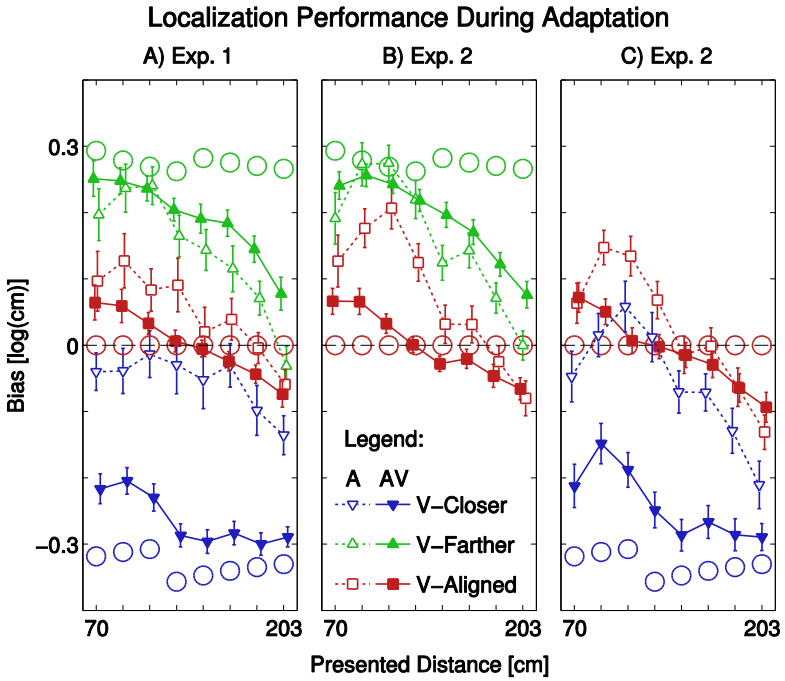

Figure 3-4 A response bias in the adaptation runs (4:8). (A) Data of Experiment 1 were taken from Figure 3-3CH. V-Aligned data were taken from run 11 Figure 3-3EJ. (B) Data of Experiment 2 were taken from Figure B-1.

3.4.1.1 Ventriloquism Effect

The analysis of response bias suggested that the observed asymmetry in between the V-Closer and V-Farther conditions could be explained by the baseline performance in the V-Aligned condition. The V-Aligned performance was assessed in runs 1 and 11. However, there seems to be systematic difference between the performance between these runs (will be analyzed further). The legitimate approach would be to average across the runs, albeit for this analysis we decided to use only runs 11 as the V-Aligned performance because it was later shown (in Experiment 2) that the performance in the V-Aligned condition during adaptation was more similar to performance in runs 11 than to the average of the runs 1 and 11 (see Sec. 3.5), and we assumed that the V-Aligned

performance during adaptation is fair perceptual baseline for our V-Misaligned conditions.

Therefor the the V-Misaligned was referenced by the V-Aligned condition with the aim to account for the discrepancies between the conditions because some of the variation could be related to the compression and bias that was introduced by the mere presence of the light, and potentially also the room learning effects as were observed in the previous chapter (Sec. 2). Since the V-Aligned can be viewed as the perceptual baseline for the V-Misaligned conditions, the magnitude of the VE should not be referenced by the physical distance of the auditory components but the perceptual baseline should be used instead. Thus Figure **3-5A** shows the VE (solid lines with full triangles) defined as a difference of the V-Misaligned *re.* V-Aligned distance localization as function of target distance (the data are taken from **Figure 3-4A**).

In addition to that, we were interested to see whether the VE could be inferred only from the V-Aligned performance. If the measured V-Aligned performance provides a perceptual baseline, i.e. the perceptual zero for the VE, it can also provide an estimate where perceptual performance would be if the the sound and light were paired at the distance of the visual adaptors that were used to induce the VE. It is possible that if the integration of the auditory and visual information was complete then the observed VE could reach the 100% of the VE; however, the decrease of the integration due to the misaligned presentation (Chan et al. 2012b) and the VE would be then expressed as the portion of the complete VE. Therefore given that the auditory distance perception can be described by the power relationship of the perceived and presented distance (Zahorik et al. 2005; Anderson and Zahorik 2014) the predictions of the complete (100%) VE were obtained from the performance in the V-Aligned conditions (black dashed lines with the corresponding open triangles on Figure **3-5A**). The details of computation of the model can be found in the methods section (Sec. 3.3.4).

The bar graphs at the bottom (D) of the figure show the across target average (±SEM) of the VE and the 100% VE. The percentage expresses the ratio of the two mean magnitudes.

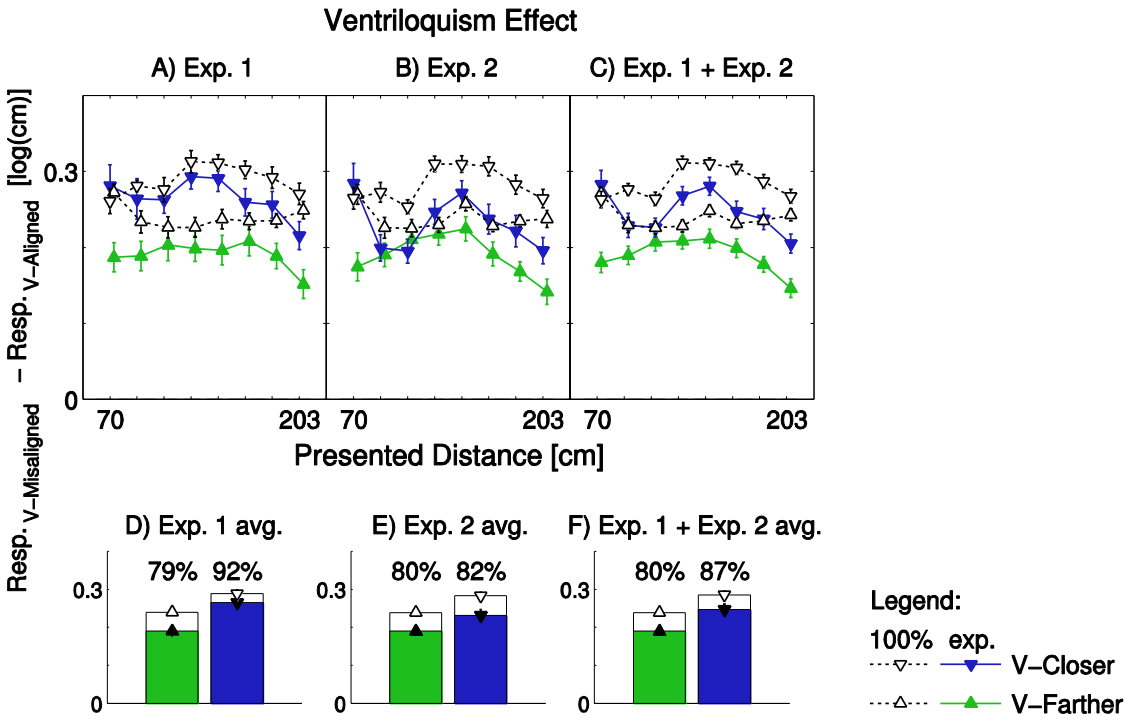

Figure 3-5 Ventriloquism effect (VE) expresses change in perceived location of the auditory targets in the distance dimension due to the presence of the visual component. Localization performance in the V-Misaligned conditions was referenced by the performance in the V-Aligned conditions. X-axis shows the target distance. Y-axis shows the difference of perceived distance of target in the V-Aligned re. V-Closer (solid blue line with closed symbols) and the V-Aligned re. V-Farther (solid green line with closed symbols) in logarithmic units. The V-Misaligned and V-Aligned data were taken from runs 4-8 form (B) Experiment 1 (A) used the V-Aligned data from runs 11. Dotted lines with open symbols show theoretical magnitude of 100% ventriloquism (C) Combined data set of the Experiment 1 and Experiment 2. Bar graphs at the bottom (D-F) show the across-target mean of VE (color bars), across-target mean of 100% VE (empty bars), and the percentage of the means.

Figure **3-5A** shows that V-Closer produced higher VE than the V-Farther ($F(1,33)=27.10$, $p<0.01$) and that the VE magnitude varied with the target distance ($F(7,231)=5.13$, $p<0.01$). The interaction of the two factors did not reach significance. The VE in the V-Closer was higher than the VE in V-Farther condition. The VE decreases with distance and the data also shown marginal variation between the conditions, i.e., V-Closer has tendency to peak in the middle while V-Farther plateaus.

The 100% VE was generally above the experimentally measured VE although the magnitudes of the 100% VE seems to parallel the experimentally measured data although the model does not fit well with the beginning and the end of the response range, which is more pronounced in the V-Farther condition. The model and experimental data seems to differ between the conditions. The V-Closer experimental data are very close to the predictions of the model, while V-Farther data deviate more. However, both could be explained as a portion of the complete VE, even with small discrepancies.

The difference of the experimental VE and the modeled VE (difference of the black dashed and color solid lines with corresponding symbols) was subjected to the RM ANOVA. The statistical analysis showed main effect of the target distance ($F(7,231)=4.23$, $p<0.05$) and the interaction of target distance and condition ($F(7,231)=3.94$, $p<0.01$).

Taken together, these results indicate the V-Aligned baseline could explain some variation in the mean judgments of auditory distance in the V-Misaligned conditions; however, the significant differences between the conditions and the variation of the magnitude with respect to target distance were preserved. Therefore the baseline could not explain all the experimental variation in the AV data. On the other hand, it does not interact with the assumption that the auditory and visual components were integrated according to their perceptual variability (Alais and Burr 2004) that increased with distance (Kopčo et al. 2012). The magnitudes of the 100% VE are parallel to the experimental data therefore it is likely that the small disparities between the positions of the LEDs (the actual disparity was not always exactly 30% but it slightly deviated due to the linear spacing of LEDs) influenced the modeled data and the small discrepancies could have also influenced the experimental data. The fact that the 100% VE are higher than the experimental data can relate the difficulty of the subjects fuse the auditory and visual components when they are presented with discrepant distances (Chan et al. 2012b). The comparison of the modeled VE and the experimental data were done by the subtraction of the modeled and experimental data. Subtraction on the logarithmic scale is parallel to multiplication on the linear space thus the comparison cannot be influenced by a constant in one or the other estimate. Despite that the fact that that the comparison showed a significant interaction of the condition and target distance possibly relates to a deviation of auditory perception at the far distances, the effect known as horizon effect

(Zahorik et al. 2005) (i.e., the perception the auditory components behind the response range deviated from the prediction of the power-model fit).

3.4.1.2 Immediate Ventriloquism Aftereffect

To express the persistence of the perceptual shifts induced by the VE, **Figure 3-6A** plots the A trials during adaptation. Analogically to the VE, the V-Misaligned response bias in the A trials was referenced by the V-Aligned response bias in the A trials. Open symbols indicate the measured magnitude of the VAE in the V-Closer (blue downward-pointing-triangles) and V-Farter (green upward-pointing-triangles).

The V-Closer and V-Farther produced similar magnitudes of the VAE. The magnitudes of the VAE slightly decreased with the target distance ($F(7,231)=4.57$, $p<0.01$), which relates to the fact that the A responses during adaptation in the V-Farther condition were more compressed than the V-Closer condition and the V-Aligned baseline (compare dashed lines of **Figure 3-4A**). These results show that the misaligned AV presentation created a perceptual bias that lasted seconds after the AV training. The magnitude of the bias was about 40%-50% of the VE (Kopčo et al. 2009). One reason why the VAE magnitudes did not vary with the conditions, similar as the VE, is that the VAE was of lower magnitude than VE while the across-subject variability was constant.

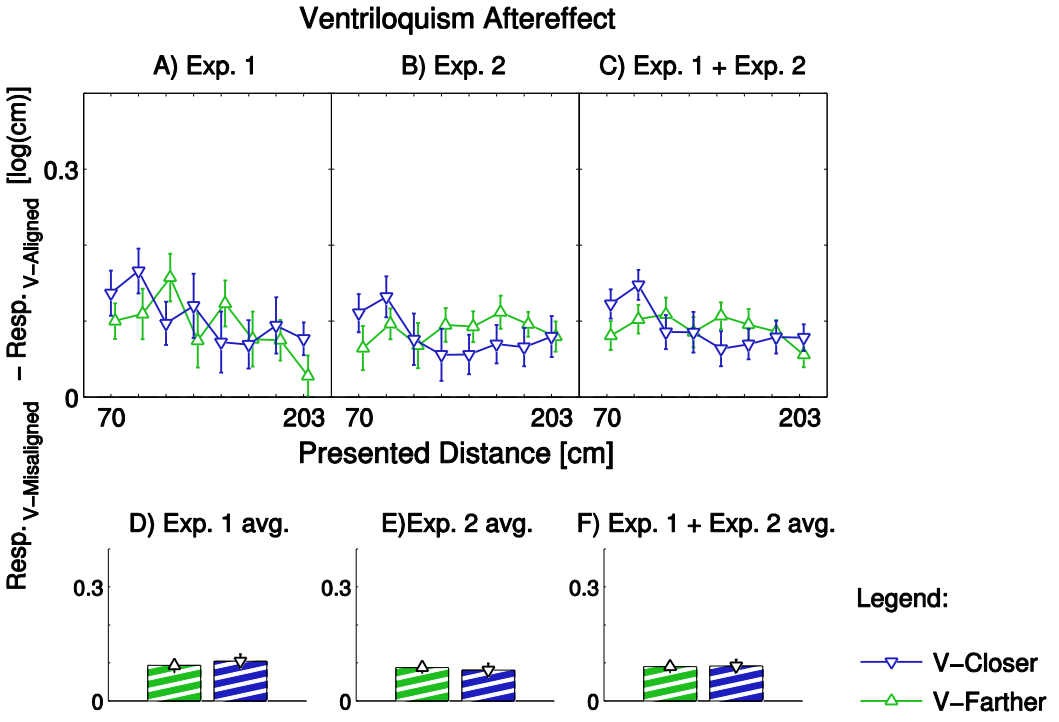

Figure 3-6 Ventriloquism aftereffect (VAE) expresses immediate persistence of auditory space shift due to VE, measured in the A trials. X-axis shows the distance

of the auditory targets. Y-axis shows the mean response in the A trials in adaptation runs (4-8) in V-Closer re. V-Aligned condition (dashed blue line with open symbols) and V-Farther re. V-Aligned conditions (dashed green line with open symbols) in logarithmic units. VAE is shown for Experiment 1 (A), Experiment 2 (B), and combined dataset (C). In-line graphics show across-target mean of the VAE. The V-Aligned data in Experiment 1 were taken from runs 11, in Experiment 2 from adaptation runs (4-8). The bar graphs at the bottom (D-F) show across target mean (± SEM) of the VAE.

3.4.1.3 Persistent Ventriloquism Aftereffect

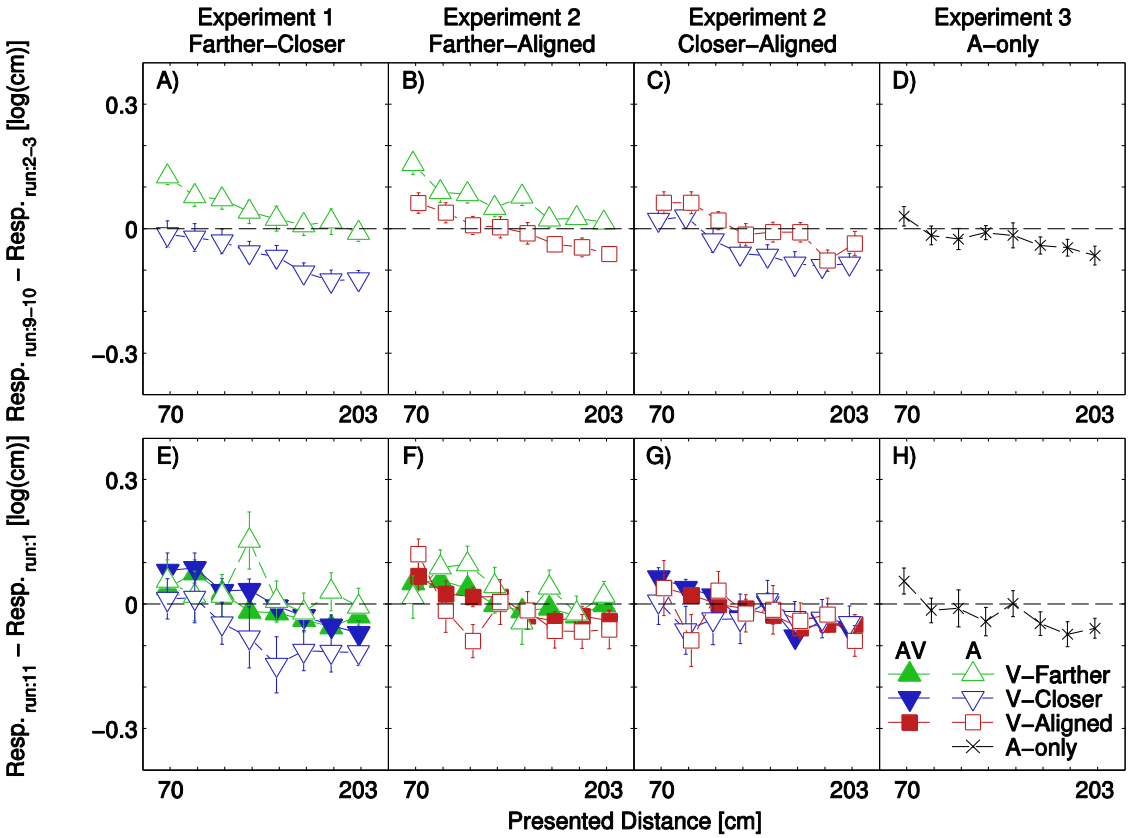

Figure 3-7 Localization compression after the adaptation period in Experiment 1, Experiment 2, and in control Experiment 3. The shift of auditory space induced by the AV presentation persists in the runs that follow the V-Aligned (squares), V-Farther (upward pointing triangles), and V-Closer (downward pointing triangles) adaptation. X-axis shows target distance. (A-D) Y-axes show the magnitude of the mean perceived distance in post adaptation runs (9-10) re. pre-adaptation runs (2-3) and (E-H) Y-axes show the final run 11 re. the initial run 1. Open symbols are

used for responses in the A trials, closed symbols for the AV trials. Data are shown for Experiment 1 (A,E), Experiment 2(B,C,F,G), and Experiment 3 (D,H).

To investigate the effect of the AV training on the post-adaptation runs, the following analysis computes how persistent the VAE was. The adaptation runs were immediately preceded and followed by the test runs that tested auditory distance localization without the visual component. The extent of the VAE could also be investigated in the final run which involved the V-Aligned condition, similarly as the first run. Therefore the following analysis, computes the post-pre adaptation contrast, and final-initial contrast, which evaluates the amount of VAE that persisted after the discrepant AV training.

Figure 3-7 shows the change in the response bias as the post-pre contrast (A) and the final-initial contrast (E). The differences of the A-only runs (A) and the V-Aligned runs (E) are shown as a function of target distance. (A) Y-axis in the upper row shows a difference of the mean response in the post-adaptation (runs 9-10) *re.* pre-adaptation (runs 2-3) when the adaptation condition was V-Farther (green dashed lines with open upward-pointing triangles) or V-Closer (blue dashed lines with open downward-pointing triangles) conditions. (E) Y-axis in the bottom row shows the response difference of the final run (11) *re.* first run (1) using the same symbols as in the upper row. The first and final runs included also the AV trials. Open symbols indicate the response bias in the A trials and full symbols indicate the response bias in the AV trials. Data of Experiment 3 (D, H) will be analyzed in analyzed in Sec. 3.6.1

The post-pre contrast (A) shows the compression of responses and the response bias in the direction of the AV training. RM ANOVA of these data with factors condition and target distance showed main effects of condition ($F(1,33)=18.44$, $p<0.01$) and the main effect of target distance ($F(7,231)=7.84$, $p<0.01$).

The final-initial contrast (E) also showed the compression in the A trials (open symbols) but did not show bias of the two conditions. The RM ANOVA showed the the main effect of target distance ($F(1,33)=8.39$, $p<0.01$). The compression was seen also in the the AV as confirmed by the the main effect of the target distance ($F(7,231)=7.59$, $p<0.01$).

These results show that the perceptual shifts persisted minutes after the audio-visual training. The responses became more compressed and biased in the expected direction.

The compression is evident between the final and first run but also between post and pre-adaptation runs, thus the compression most likely took place during the adaptation runs.

3.4.2 Standard Deviation of Response

To understand how vision influences auditory distance perception in terms of intra-subject variability, the following analysis shows how the visual congruency influences response SD.

Figure 3-8 shows mean (+-SEM) within-subject standard deviation as a function of target distance. Data are visualized in the same format as **Figure 3-3**. The SDs in the A trials are shown with the open symbols, closed symbols represent AV trials. The adaptation data are shown in the middle column (C,H) showing the V-Closer (blue downward-pointing-triangles) and V-Farther (green upward-pointing-triangles) conditions. The surrounding columns shows data in the pre-adaptation (B,G), post-adaptation (D,I) showing the A-only runs. The first run (A,F), and final run (E,J) show V-Aligned performance. Data in each panel were pooled across runs shown above each column. Two rows represent two sessions. Since the ratio of the A and AV trials varied in the runs, the randomization procedure assured that the SDs were computed always from 6 measurements.

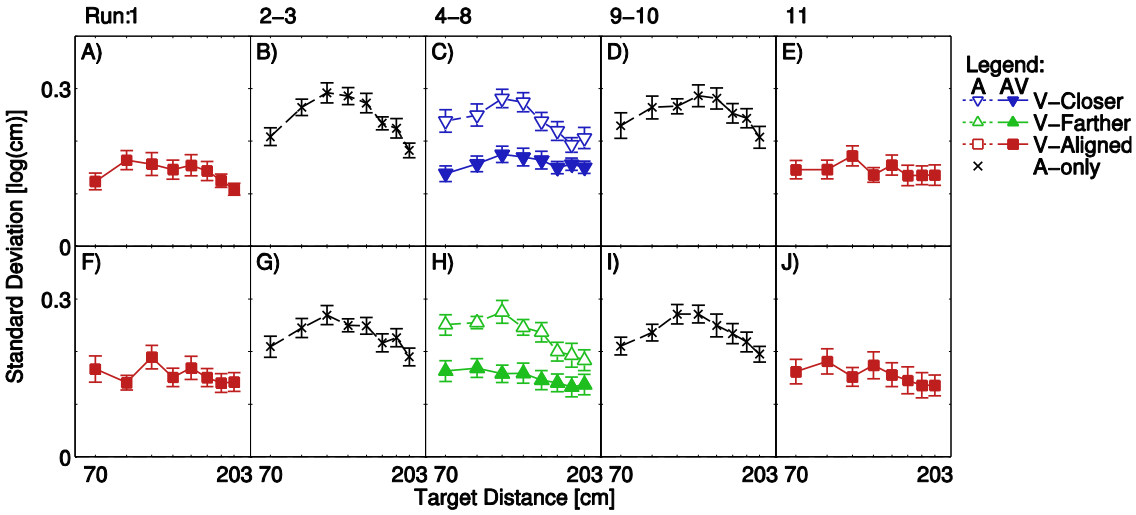

Figure 3-8 Experiment 1 across-subject standard deviations (SD). Within-subject SDs were computed separately for each target distance from equal number of measurements in each data bin. Data were pooled across runs depicted above each column. Two columns represent two sessions. Data are shown in the same format as Figure 3-3. V-Aligned data (A,E,F,J) contain only the AV data because only two measurements per target were collected for the A condition in these runs.

Overall, the AV SDs were lower than the A SDs. The SDs in the A trials varied with target distance. There is not a difference between the AV SDs in the V-Aligned and V-Misaligned conditions, as well there is no difference between the adaptation, pre-adaptation, and post-adaptation SDs, which means that the SDs did not vary as a function of experimental run.

The statistical analysis RM ANOVA with factors of target distance, trial type (A,AV), and condition (V-Closer, V-Father) of the adaptation runs (4-8) showed the main effect of trial type ($F(1,33)=110.55$, $p<0.01$), main effect of target distance ($F(7,231)=10.51$, $p<0.01$), and the interaction trial type and target distance ($F(7,231)=4.54$, $p<0.01$). No other main effects or interactions reached significance. Additional statistical analysis assessed the statistical difference in A trials between the adaptation runs vs. pooled post-adaptation and pre-adaptation runs (t-test: $p<0.05$). Another statistical test assessed the difference between the AV V-Aligned (runs 1 and 11) and AV V-Misaligned data (runs 4-8) and showed no statistical difference (t-test: $p>0.05$). The auditory-only data in V-Aligned condition did not have enough measurements per target speaker.

Evaluation of the SDs shows that the presence of the visual component increases the precision of subject's response regardless of the magnitude and direction of audio-visual disparity. SDs in the AV trials were approximately constant on the logarithmic scale, as predicted by the Webber-Fechner law although the interaction of distance and type of presentation reached significance which suggests that A SDs slightly varied with target distance. However, that could relate to the experimental apparatus. The SDs of the A trials were not affected by the presence of the interleaved AV trials.

3.5 Results: Experiment 2

Experiment 2 was identical to Experiment 1 with the exception that the subjects were assigned into one of two groups which differed only by the AV disparity during the adaptation. While all subjects in Experiment 1 underwent V-Closer and V-Farther adaptation, the subjects in Experiment 2 either underwent V-Closer and V-Aligned or V-Farther and V-Aligned training. The V-Aligned condition was introduced because we aimed to provide a solid perceptual baseline for the VE. In Experiment 1 we observed the temporal drift in the V-Aligned performance during the session, therefore it was likely that the the performance in our task was influenced by the room learning, subject fatigue, or related learning paradigm.

3.5.1 Response Bias

Since Experiment 2 was only a slight modification of Experiment 1, the response biases observed in the Experiment 2 were qualitatively similar to Experiment 1 as can be seen in panel B of **Figure 3-4** which compares the response bias of the the AV and A trial in adaptation runs across the experiments (see **Figure B-1**). However, there seems to be a systematic difference between the V-Aligned performance between the adaptation runs and runs, first run, and the final run, which was confirmed statistically. The difference of the V-Aligned performance between the runs was assessed by RM ANOVA using the data of Experiment 2 of the V-Aligned sessions. The test included factors of run (1, 4-8, 11), trial type (A, AV), and target distance. The data were averaged across adaptation runs 4-8. The test showed except the main effect of speaker, main effect of type, interaction type x speaker (were not in particular interest), also the highly significant interaction of run x speaker ($F(14,1106)=3.43$; $p<0.01$). The interaction run x type reached marginal significance ($F(2,158)=3.17$; $p<0.1$ without the correction for sphericity $p=0.045$). Therefore it seems that the V-Aligned performance changed during the session. The change was modulated by the trial type and target distance.

These results suggest that auditor distance perception was not only modulated by the visual cues but also by the experience with room (Calcagno et al. 2012).

3.5.1.1 Ventriloquism Effect

Figure 3-5B shows the magnitudes of the VE for Experiment 2. The VE was defined in the same way as for Experiment 1. It was the difference in the mean response in the AV trials in the V-Misaligned re. V-Aligned condition. However, in Experiment 2 the baseline was measured in a separate session with the V-Aligned adaptation runs (4-8). The figure also shows the predictions of the complete VE which was also obtained from the V-Aligned sessions form adaptation runs 4-8.

The results of the Experiment 2 slightly differs from the results of Experiment 1, mainly due to differences in the baseline performance although the main trends were preserved. The data of Experiment 2 vary with target distance. The difference between the V-Closer and V-Farther is slightly lower than in Experiment 1 especially at targets below 1.5 m where the V-Closer and V-Farther data are crossed. At distances above 1.5 m the lines are clearly separated although both conditions are considerably compressed in that distances. One more difference between Experiment 1 and Experiment 2 is that

the 100% VE in V-Closer condition in Experiment 2 is more separated from the experimental data than in Experiment 1 (where 100% VE is almost aligned with the experimental data).

The statistical analysis showed that the VE varies with distance, i.e. the VE magnitude peaked in the middle of the response range ($F(6,468)=12.09$, $p<0.01$), and the target distance and condition were in interaction ($F(6,468)=3.03$, $p<0.05$). The interaction can be explained by the inspection of the figure which shows the difference of the two conditions is lower at targets closer than 1 m and the difference increases towards the end of the response range. The main effect of the condition was only marginally significant ($F(1,78)=2.99$, $p<0.1$). The analysis was done without the nearest target distance because the target in the V-Closer sharply increased, possibly because the LED used in AV trials in the V-Closer condition was only 50 cm from the subject where the array of the loudspeakers actually started. Thus it was clear to the subject that the sound was not produced at that specific location. Additionally, the VE data were subtracted from the predictions of the compete VE which showed the main effect of target distance ($F(6,468)=6.09$, $p<0.01$) and the interaction of distance and condition ($F(6,468)=4.41$, $p<0.01$). The main effect of condition was not significant ($F(1,78)=1.43$, $p>0.05$).

Figure 3-5C shows the estimate of the VE on the combined dataset. The magnitudes are dominated by the Experiment 2 because it involved almost twice as much subjects as Experiment 1. The statistical test was not conducted because of the differences in the designs of the experiments (Experiment 1 – within-subject, Experiment 2 – between-subject). These data show that VE V-Closer (87% of the complete VE) is slightly higher than the VE V-Farther (80% of the complete VE) in terms of magnitudes and the proportions of the complete VE.

The VE in Experiment 2 showed the interaction of target distance and condition which was not present in Experiment 1, the main effect of condition which was in Experiment 1. The difference can be attributed to the V-Aligned baseline. The magnitudes of the VE in both experiments are similar. However, the current statistical results are more reliable since the Experiment 2 was designed with the intention to provide stable estimate of the V-Aligned performance. As in the previous experiment, the model of the complete VE explains some of the variation in data, but it cannot wholly explain the data. However, the V-Closer data seems to be predicted more accurately (i.e., the 100% VE magnitudes

parallel the experimental data) while V-Farther data deviate at the beginning and end of the target array.

3.5.1.2 Immediate Ventriloquism Aftereffect

The analysis of ventriloquism aftereffect was conducted in a similar fashion as for Experiment 1. However, now the A V-Aligned baseline performance was obtained in the adaptation runs (4-8) in a separate sessions.

Figure 3-6B shows the magnitude of the VAE defined as the difference in the mean response in the A trials of V-Misaligned re. V-Aligned condition in the adaptation runs (4-8). It evaluates the immediate persistence of the perceptual shift due to the AV training.

The results were similar across experiments although the VAE magnitude in Experiment 2 did not change with with target distance ($F(7,546)=0.84$, n.s.) as in Experiment 1. This difference between Experiments can be explained by the difference in the baseline. In Experimetn 2, there was also no difference between conditions ($F(1,78)=0.1$, n.s.), confirming the results of Experiment 1.

Figure 3-6C shows the combined data set. The estimate of the VAE is almost identical to the estimate of Experiment 2. The magnitude is about 40%-50% of the VE.

3.5.1.3 Persistent Ventriloquism Aftereffect

Experiment 1 showed that the AV training influenced the sound localization even in the runs without the visual component (9-10) and also in the final run (11), which included V-Aligned stimuli. The similar analysis was conducted on the dataset of Experiment 2.

The temporal effects discussed previously and shown in **Figure 3-7AE**, were confirmed in Experiment 2 (**Figure 3-7BCFG**), i.e. the A trials were biased and compressed in the pre-post contrast and the AV were compressed in the final-initial contrast, even if the A trials did not show the compression (main effect of target distance: $F(1,78)=2.24$, not significant) as was the case in Experiment 1. These trends were confirmed statistically, data are not shown. Additional RM ANOVAE for Experiment 2 was conducted on the A data in the post-pre contrast such that V-Closer and V-Farther data were referenced by V-Aligned condition and adjusted for the direction of the AV disparity (contrast of the red line and color lines on **Figure 3-7BC**). The statistical analysis showed no significant effects, i.e., (B-C) blue and green lines are equally separated from the red lines. It means that the direction of the induced shift did not

influence the persistence of response bias in the post-adaptation runs, which is consistent with the immediate VA.

3.5.2 Standard Deviation of Response

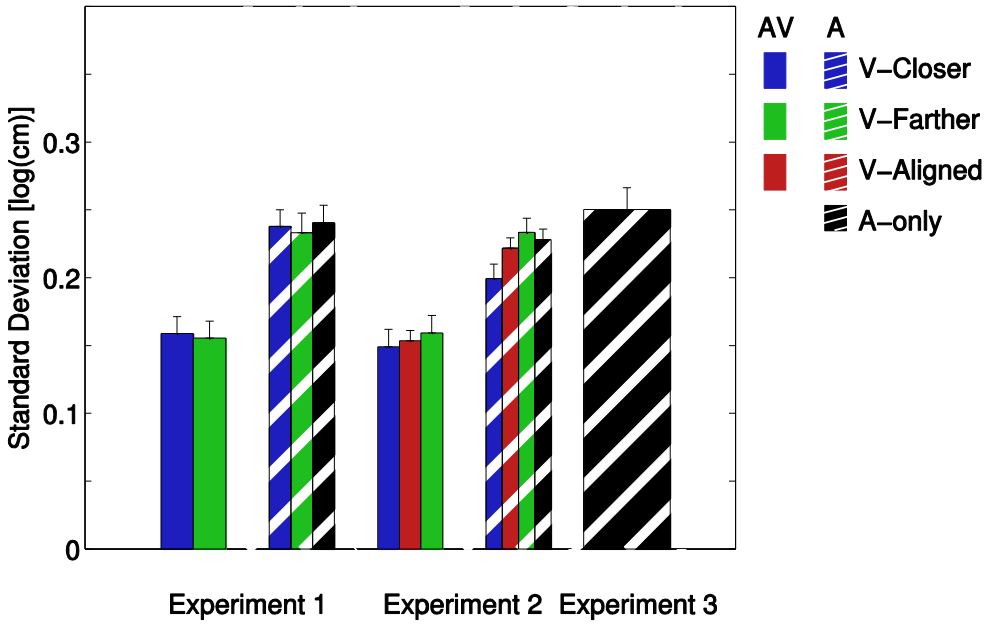

Figure 3-9 Summary of response SD in the A (hatched) and AV (full) trials averaged across target locations. Data are shown for adaptation runs (4-8) (color – see legend; however Experiment 3 was in black), and A-only runs runs (2,3,9,10) (black). Data were pooled across corresponding runs and the SDs were computed separately for each target (6 independent measurements). In Experiment 1, all conditions were performed by all subjects, the V-Aligned data are not shown because they were collected in runs 1 and 11 and the figure aims to compare AV performance in adaptation runs. In Experiment 2, V-Closer and V-Farther were conducted by independent groups. The V-Aligned data were pooled across these groups. Experiment 3 shows A-only performance in runs 4-8 (see Sec. 3.6.1).

Experiment 1 showed that the presence of the visual component decreases response SD when the sounds were localized in distance dimension. It also showed that this decrease of response SD was unaffected by the congruency nor by the direction of congruency. These results were also observed in Experiment 2 (see **Figure B-2**). However, the V-Aligned data in Experiment 1 did not measure performance in the A trials and these data were collected in run 1 and 11, which could have been influenced by the room learning.

Figure 3-9 summarizes the response SD in Experiment 1, Experiment 2, and Experiment 3 (analyzed in Sec. 3.6.1) in A (hatched bars) and AV (full bars) trials. Data were obtained from adaptation runs (4-8) (shown in color except Experiment 3 which is shown in black), and A-only runs (2,3,9,10) which were shown in black. The figure shows that the trends in Experiment 2 were also observed in Experiment 1. The AV trials had higher response SD than the A trials, the response SD in AV trials was unaffected by the direction of disparity, and the AV V-Aligned data had similar response SD as the AV V-Misaligned data. However, in data of Experiment 2 differed to data of Experiment 1 in one peculiar aspect. The response SD in the A V-Farther were higher than the SDs in the A V-Closer.

First, a statistical analysis on data of adaptation runs was conducted without the V-Aligned data (blue and green bars of Experiment 2). RM ANOVA with factors of trial type, target distance, and condition showed a main effect of trial type ($F(1,78)=117.04$, $p<0.01$), target distance ($F(7,546)=18.61$, $p<0.01$), interaction of the trial type x target distance ($F(7,546)=12.84$, $p<0.01$), and interaction of trial type x condition ($F(1,78)=4.7$, $p<0.05$). The V-Aligned condition was compared to the V-Farther (A: t-test: $p>0.05$; AV: t-test: $p>0.05$,), and V-Closer (A: t-test: $p>0.05$; AV: t-test: $p>0.05$). The statistics did not show any difference between the means.

These results show that the presence of visual component decreased response SD regardless whether it was presented in the same distance or it was presented from different distance. The results also confirmed the findings of Experiment 1 that the the SDs in the A trials varied with distance, while the SDs in the AV trials were generally constant. These results are in line with the previous Experiment 1. However, an unexpected interaction of the condition and the trial type showed that the SDs in the V-Farther were lower than the V-Closer in the A trials. This result does not have a trivial explanation especially because the interaction was not present in Experiment 1, and this difference was not observed in the AV trials. Nevertheless, the methodological difference between the experiments could have influenced this result.

3.6 Results: Experiment 3 and Experiment 4

3.6.1 Experiment 3: Auditory-Only Experiment

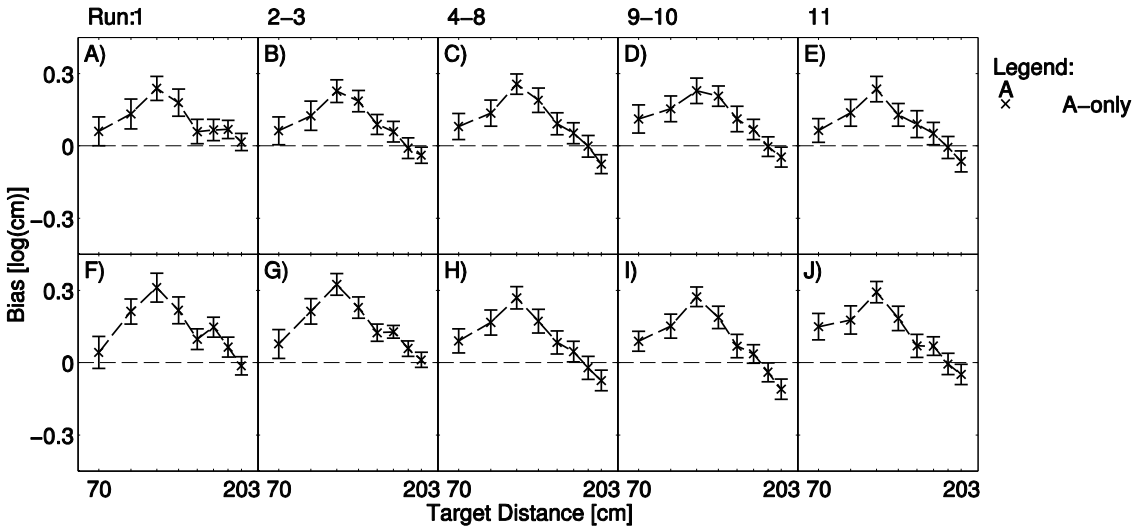

Figure 3-10 Experiment 3 response bias in the A-only condition as a function of target distance. The figure layout is identical with the layout of Figure 3-4. The rows represent sessions, and columns divide the experiment according to the identical scheme as was used in the Experiment 1, (initial run, pre-adaptation, adaptation, post-adaptation, final run); however, in this experiment subjects did not receive AV training.

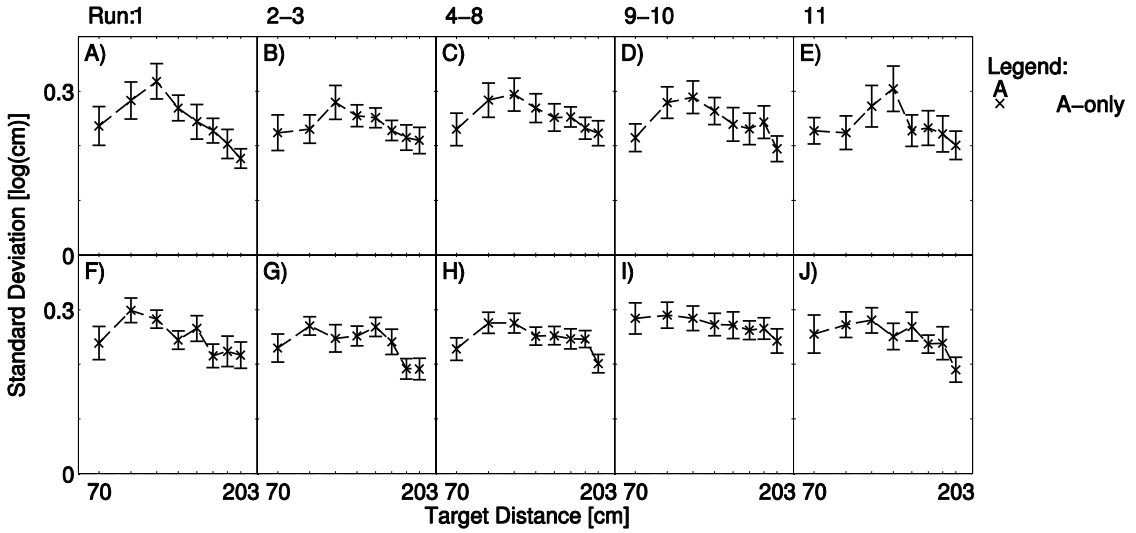

Figure 3-11 Experiment 3 response SD. The data were computed in the identical way as in Experiment 1 (from 6 independent measurements). The layout of this figure is identical to layout of Figure 3-8. Rows represent sessions, columns divide the experimental sessions according to the scheme as was used in the Experiment 1.

Localization bias **Figure 3-10** and response standard deviations **Figure 3-11** are shown for the con trol Experiment 3. Since the procedures of this experiment were exactly

same as in Experiment 1 the figures are shown with the same format as the **Figure 3-4** and **Figure 3-8**, respectively. The rows represent the two sessions, the columns show the runs in the same scheme as was used in Experiment 1. However, in this experiment all runs were conducted with the identical A-only condition.

These data show that subjects mostly overshoot the true distances, overall the localization performance is similar across experiments. The power model ($d' = kd^a$) fits for the mean data (across all conditions and subjects) reached a = 0.83 and k=2.12 .

In Experiments 1 and 2, the responses became more compressed during the session. **Figure 3-7DH** shows that similar compression was observed also in the control experiment, therefore it is likely the compression observed in the main experiments can be explained (to certain extent) by the adaptation to room reverberation.

Figure 3-9 shows the mean intra-subject response SD across all experiments. Across-subject mean SDs of Experiment 3 computed in runs 4:8 were compared to the SDs in Experiment 1 (Welch's t-test: p>0.05), and Experiment 2 (Welch's t-test: p>0.05). The statistics did not show a significant difference between the experiments.

3.6.2 Experiment 4: Visual-Only Experiment

Experiment 4 was an experiment that provided estimates of the visual distance perception of the LEDs that were used in the AV training. The performance was assessed using the power-model fit. **Figure 3-12** shows across-subject mean judgments and the power model fit on the mean judgments. This model fit on averaged data was used in model in the following chapter (Sec. 4) to estimate the visual percept. The within-subject SD was computed as the SD of the error of the power model computed separately for each subject.

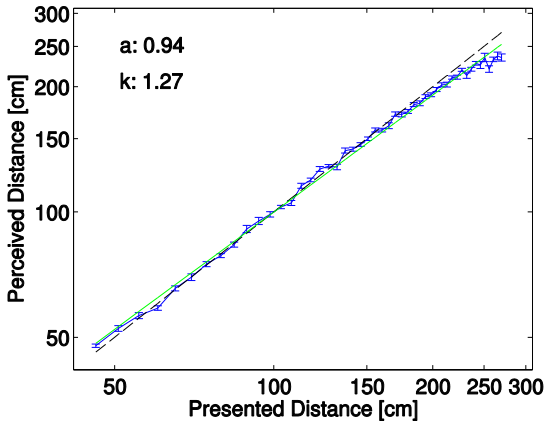

Figure 3-12 Mean visual distance judgments (±SEM) (blue line) as a function of target distance. The figure shows also the parameters of the power model fit $d' = kd^a$ (green line) on the average responses. Black dashed line shows the reference.

3.7 Discussion

The study investigated the effect of visual stimulation on auditory distance perception in a localization task, in which the target noise bursts were paired with the flashes of LEDs. The relative AV disparity was fixed so that the V component was either 30% closer or farther than the distance of the auditory target. The subjects experienced so called 'ventriloquism effect' – the location of the sound was perceived near the position of the visual adaptor. In the V-Closer condition, the sounds were always perceived in proximity of the visual components while in the V-Farther condition, the perceived distance of the sound did not follow the visual component and decreased at distances greater than 1.5 m. These observations can relate to audio-visual integration scheme that combines the auditory and visual information according on the reliability of the underlying cues (Alais Burr 2004) and the decrease of localization blur with increasing distance. Since the decrease of blur was larger in V-Farther condition than in the V-Closer condition the biases, the V-Farther adaptors attracted the auditory targets to lesser extent. Although these results support the previous observations of the asymmetry in processing of the closer vs. farther visual adaptors (Zahorik 2003; Mershon et al. 1980) the response compression in the V-Aligned baseline explained much of the variation between the V-Closer and V-Farther conditions. Even after accounting for the shifted baseline, the VE magnitudes significantly differed between the V-Closer and V-Farther conditions most notably at distances above 1.5 m. These observations were compared to a model of the complete VE i.e., how subject's responses to discrepant AV stimuli would look like if the auditory and visual components originated from the distance of the visual component (V-Aligned condition at the distance of visual adaptors). The V-Closer data reached 87% and V-Farther data reached 80% of the predicted values, which suggests that the magnitudes of the VE were influenced by the actual AV disparity. In addition, the results showed that the perceptual shifts persisted on the interleaved trials and in the post-adaptation period, which for the first time demonstrated the VAE in the distance dimension. The magnitudes of the VAE were equal in both directions of the AV disparity, which reached approximately 40-50% of the VE magnitude.

The AV disparity was induced on the percent scale i.e., the physical disparity was fixed in linear units but in logarithmic space the V-Closer disparity was higher than V-Farther disparity. This difference can explain the difference in magnitudes of the VE of the V-Closer and V-Farther conditions. Given that, it is unexpected that the magnitude of the VA was unaffected by the direction of the induced shift. The reason can either relate to (a) high across subject variability (our design was not sensitive enough to detect the small disparity induced by VE), (b) the transfer on the neural from the V-Farther is stronger than in V-Closer, (c) the V-Closer and V-Farther were equal on the neural level, or (d) it can relate to a difference on the response level.

The early reports of the audio-visual integration in distance dimension (Gardner 1968; Mershon et al. 1980) came up with the hypothesis that the visual component dominated the perception of auditory distance. In the current study, the distance judgments were rather a mixture of visual and auditory components as was shown in the previous experiments (Zahorik 2001; Calcagno et al. 2012).

The observations of the bias in the A trials during the adaptation period, which showed that the V-Farther biases were closer to the visual adaptors than the A responses in the V-Closer condition, seems to be in opposite direction than Min and Mershon's (2005) data. The study was measuring the visual capture in depth using the adjacency principle and suggested that adaptors that were placed in front of the auditory target induced slightly higher bias than the adaptors behind. In our data we observed higher response bias in the A V-Farther during adaptation; however, the magnitude of the bias in V-Farther decreased with increasing reference distance and V-Closer slightly increased (get closer to V adaptors), which can explain the difference. Moreover, the current study directly assessed the VAE while the VAE Min and Mershon's (2005) could be inferred only indirectly. In addition to that, the assessment of the VAE with the V-Aligned reference showed no difference between the conditions, which is closer to the Min and Mershon's (2005) observations.

The estimates of the magnitude of the VAE is close to the estimate of Kopčo et al. (2009), and in between of the estimates of Bertelson et al. (2006) who observed 30% and 80% observed by Recanzone (1998). However, the previous studies were performed in horizontal plane thus further studies and models are needed to unify these estimates.

Our results also showed that the responses became more compressed during the course of the experimental session even when the sounds were accompanied with the

aligned visual adaptors. The compression was evident in the AV responses from the initial to the final run and during the adaptation runs. These results can relate to fatigue, or attentional factors although it was shown that auditory distance perception is influenced by the experience with the room acoustics (Coleman 1962; Mershon 1989). Another study (Calcagno et al. 2012) also showed that auditory distance perception was influenced when the subjects were first allowed to walk in the room and familiarize themselves with the dimensions of the room. Potentially, the result of the current study can relate to the quick adaptation to room acoustics in observed in the Experiment 3 or the experiments in the previous chapter, which were done without the AV training. In the context of the Bayesian inference models, the result can be understood as a change of the prior distribution (Wozny and Shams 2011a). The prior distribution affects the statistics of the sensorial output as in opposed to the process of cue extraction. In the current study, the subjects could become familiar with the response range and directed all the responses towards middle.

Our results also showed that presenting the sound accompanied with the visual stimulus increases the precision in terms of response standard deviation, which is in line with the observations of Zahorik (2001) and Anderson and Zahorik (2014). However, it is in contrast with the findings of Calcagno et al. (2012) who did not observe the increase of response standard deviation. In Anderson and Zahorik (2014) and the current study, the visual and auditory stimuli were turned on and off simultaneously and it is known that temporal synchrony enhances the perceptual fusion (Recanzone 2009) which may explain the difference with Calcagno et al. (2012) where the scene was lightened with the LEDs for the whole course of the experiment. Although Zahorik (2001) observed also decrease of response standard deviation in 'vision' condition without the synchronous presentation, the subjects in his study could use more visual cues than in Calcagno et al. (2012). Response SDs were unaffected by the AV shift direction, which suggests that the stimuli in various conditions were perceived equally reliably.

The analysis also sought to answer the question whether the decrease of response AV SD transfers to the A trials, i.e. whether auditory spatial perception can benefit from the visual information not necessarily paired with the sound. In Experiment 1, we did not observe any sign of such influence because the response SD in the A trials were similar across conditions and across run types (i.e., A SDs in A-only runs were similar as A SDs in adaptation runs) . However, in Experiment 2 in analysis of data of adaptation runs we

observed a significant interaction of the run type and condition, which suggested that the decrease of SD in AV trials transferred to the A trials but only in the V-Closer condition. Since this was an unexpected observation and it was not observed in Experiment 1, the only rationale can relate to the methodological differences between the experiments. Before drawing any conclusions this results should be confirmed by the future experiments.

The increase in the response SD can be predicted by the audio-visual integration schemes that were studied in the horizontal plane (Alais and Burr 2004) and in other cross-modal integration paradigms e.g., slant perception (Ernst and Banks 2002). The increase of precision (decrease of perceptual variance) is considered to result from the theory of optimal decision, i.e. the information from the two modalities are linearly weighted with the weights proportional to variances of individual modalities. The theory is compatible with Maximum Likelihood Estimator (MLE) which can be extended by the Bayesian observer theory (Landy et al. 2011). The theory predicts that the variance of the combined percept will be lower than the variance of individual modalities. The perception of the visual stimuli in the current setup was assessed independently (Sec. 3.6.2) and showed that the V SDs are lower than the AV SDs. Anderson and Zahorik (2014) observed, similarly to the current results, that their correlation coefficients also do not follow the predictions of optimal integration. In Anderson and Zahorik (2014), response consistency (r^2) in the AV presentation was lower than the r^2 of the V presentation (the optimal integration predicts the opposite). Such result can be explained in the framework of the Bayesian statistics only if the likelihood function is not normally distributed (Knill 2007; Seydell et al. 2011) or if the auditory cues are up-weighted (Rosas et al. 2007, 2005; Oruç et al. 2003). Nevertheless, neither the current study nor the results of Anderson and Zahorik (2014) could be used to estimate the true perceptual sensitivity because the employed response method was not sufficiently sensitive. The subject judgments are likely to be cofounded with instructions, decision noise, or attentional factors therefore the observed SDs does not necessarily reflect the perceptual sensitivity. On the other hand these considerations do not disqualifies our findings that the AV SDs were lower than the A SDs nor the fact that the direction of the induced shift did not affect the response SD. Notwithstanding, the interpretation must be careful in generalization of these findings.

Finally, the mechanism that brain uses to merge various sources of information in distance seems to operate on the space of perceptual representation of the visual and

auditory modality. Current results suggest that the information from the visual and auditory modality were not combined optimally. The increase of compression in the subject responses is likely to represent a different process similar to an adaptation to the current distribution of stimuli.

4 Model of Audio-Visual Integration in Distance

4.1 Abstract

The data of the audio visual experiments (Sec. 3) were fit to the Bayesian model of the AV integration (Bresciani et al. 2006) with the coupling prior represented as the Gaussian ridge on the diagonal that allowed the coupling of the auditory and visual components. That in result could decrease the weight of the visual component and explain the difference between the behavioral data and the predictions of Maximum Likelihood Estimator (MLE) model. The Bayesian model was successful in explaining the behavioral data. The model explained the 88% in Experiment 1 (Sec. 3.4) and 75% in Experiment 2 (Sec. 3.5) of the experimental variance. Our data provide further evidence that the perceptual integration follows the linear weighted combination model, albeit in our experiments the weights of the audio-visual stimuli in distance dimension did not follow the optimal combination rule.

4.2 Background

The brain characterizes the objects in the outer world using the information from the senses and infers the most likely explanation of the scene (Trommershäuser et al. 2011). For example, when we hear a bird song in a garden, it is likely produced by the blue jay sitting on the fence. Although that is the most probable explanation, our percepts could be corrupted by the noise from a radio and instead of hearing the blue jay, even if we see it, we could hear someone calling us to the dinner. The brain has an innate function to interpret the scene such that it takes into account the whole context in order to keep the stable percept, yet the current sensorial readings can interact with each other. However, there are two possible outcomes of the interaction, either the current sensorial reading was internally explained as a single event – the blue jay was singing a bird song (what we saw and heard came from the position of the blue jay) or there were more events that lead to the current sensorial reading – the blue jay was sitting on the fence (what we saw) and someone was calling us to the dinner (what we heard). Furthermore, the brain has a natural tendency to integrate the multimodal stimuli that come at the same time and involve the disparity. The integration causes the perceptual bias. However, as the disparity increases, the integration falls apart, which leads to higher perceptual bias.

The previous studies (Jack and Thurlow 1973; Radeau and Bertelson 1977; Bertelson et al. 2000; Kopčo et al. 2009; Shinn-Cunningham et al. 1998) that were using

bimodal stimuli i.e., presenting a stimulus from one modality that was temporally related to a stimulus from a different modality, often observed a perceptual bias even if the the stimulus from the other modality was to be ignored. It means that either the subjects in that studies were not ignoring the other stimulus, or, more parsimonious, that the stimuli interacted on the lower than cognitive level and the stimuli interacted perceptually prior they were available for the cognitive decision. Such sensorial interaction is often described by the by an assumption that the sensory information is combined by a linear weighted combination (Landy et al. 1995; Zahorik 2002b; Alais and Burr 2004; Ernst and Banks 2002):

$$s_w = \sum_i w_i s_i \tag{1}$$

in which s_i is the sensorial reading of the i-th modality and w_i is the weight of each modality, s_w is the resulting output. In this model, the weight of 1 means the complete dominance (winner-takes-all strategy) while the other modalities are suppressed. Such an example was reported in an experiment of the auditory distance perception in anechoic room (Gardner 1968). When the dummy speaker was placed in front the real target, the sound seemed to originate from the visual target, the audio-visual percept was dominated by the visual modality, however, when the subjects were allowed to move their percept changed because they could use more information about the scene.

A different strategy of perceptual weighting was observed when the percept of the single modality was only partially influenced by the other modality. For instance, in an auditory distance experiment (Zahorik 2001), it was shown that visual component influences the position the auditory target such that perceived position was not dominated. The auditory component was biased and response standard deviation decreased in the presence of light. Possible explanation is that the visual component was more reliable than the auditory component, therefore the visual component attracted the perceived position more than the auditory component as well as it decreased the standard deviation of response to the auditory component which suggested that the perceptual weighting was influenced by the relative reliability of the underlying sensorial inputs.

The perceptual weighting (**(1)**) explains the example of visual dominance in auditory distance in anechoic space (Gardner 1968) as well as the example of combined weighting in the reverberant room (Zahorik 2001). In the anechoic space (Gardner 1968) the main auditory distance cue is sound level. If one does not a priori know how loud the sound

source should be then the loudness does not provide a cue and the subjects did not have any other information about the distance of the sound. On the other hand the visual information provided salient information and the subjects had reasons to expect that the sound could be originating from the dummy loudspeaker, therefore they perceived the sound coming from the dummy. In contrast in reverberant room (Zahorik 2001), the information about distance is more reliable therefore the visual component did not completely dominate and the percept shifted toward the auditory component. Therefore the perceptual weights that subjects used in these two examples were related to how reliable the information about distance was.

Despite these two specific studies did not evaluate the perceptual weighting, nor it could be assumed that the two sources of information were perceived as a unified events, under near ideal conditions (Alais and Burr 2004) in which it is reasonable to assume that the bimodal audio-visual stimuli are coregistered i.e., perceived as a single event despite the spatial disparity between two stimuli, it was shown that the perceptual weight w_i follows the optimal integration rule, i.e., the perceptual weight is a weighted combination of the reliabilities $r_i = 1/\sigma_i^2$ (σ_i^2 is variance of the i-th input) of the sensorial inputs.

$$w_i = r_i / \sum_i r_i \tag{2}$$

Therefore the inputs with higher reliability induce higher bias as compared to the sensors with lower reliability. The weighting of this model (**(2)**) provides an unbiased estimate of the combined percept if it is reasonable to assume that the individual cues are independent and unbiased. This estimator can be seen as the Maximum Likelihood Estimator[1] (MLE) (Cochran 1937) because the weights determine equal to the best estimate of the statistical model in which the statistically independent sensorial inputs are normally distributed with mean μ_i and standard deviation σ_i. Further, it can be seen that the weights in this model will always sum up to 1 because each reliability r_i is scaled by the total reliability.

$$r_w = \sum_i r_i \tag{3}$$

[1] MLE also stands for more general procedure of finding the optimal parameters of statistical models with arbitrary structure, in this work MLE model refers to a statistical model with non-informative prior and likelihood function with Gaussian bivariate random variable without correlation.

The MLE model is also an optimal integrator. Adding more and more statistically independent sources reduces the variability of the result i.e., the resulting reliability cannot be lower than any of the individual inputs. Therefore in the ideal conditions, in which all information relate to one event, the subject should integrate all available information and always increase reliability; however, if the stimuli contain disparity the brain does not need to explain it as a single event and two events would interact with each other. Even more the decision can be influenced with the strategy which is however a different process (Wozny and Shams 2011a). To illustrate it, **Figure 4-1** shows an example of the optimal integration (**(2)**) of the AV stimulus applied to the data from the audio-visual integration experiment in the previous chapter (Sec. 3). The figure shows the representation of the auditory target (solid line) at distance of 153 cm and visual component (dashed line) at distance of 90 cm (condition V-Closer) plus the combination according to the MLE model. As the input it takes the estimates of the distance and response standard deviation of the auditory component from Experiment 3 (Sec. 3.6.1) and visual distance component form Experiment 4 (Sec. 3.6.2). The figure shows that the combined distribution is almost dominated by the visual component and the resulting variance is lower than the estimate of the visual component when presented in isolation, as predicted by the **(3)**. However, the subject data (red line on **Figure 4-1**) show larger difference between the visual component and the optimally combined prediction. Thus it is more likely that the visual component was actually down-weighted with respect to the auditory component. In the context of optimal integration that would mean that actual variance of the visual component was increased, which was also mentioned in the similar study (Anderson and Zahorik 2014) that was testing auditory distance perception such that the auditory stimuli were paired with the congruent visual components. The study observed that r^2 values of the AV condition did not exceed the V-only condition. Also in our AV experiments the response standard deviation in the AV trials (approximately 0.17 log(cm), **Figure 3-9**) is much higher than the estimate of the standard deviation of the the visual component when it was presented in isolation (0.09) (see Sec. 3.6.2). The task in our experiment was to indicate the perceived auditory distance while ignoring the visual target and it seems that the brain uses the mechanism to automatically combined these percepts even if the percepts were completely fused. The automatic combination of the multimodal percepts was observed in previous study and one way of modeling the incomplete fusion is to allow the coupling between the auditory and visual components in the framework of the Bayesian modeling (Seydell et al. 2011; Battaglia et al. 2003).

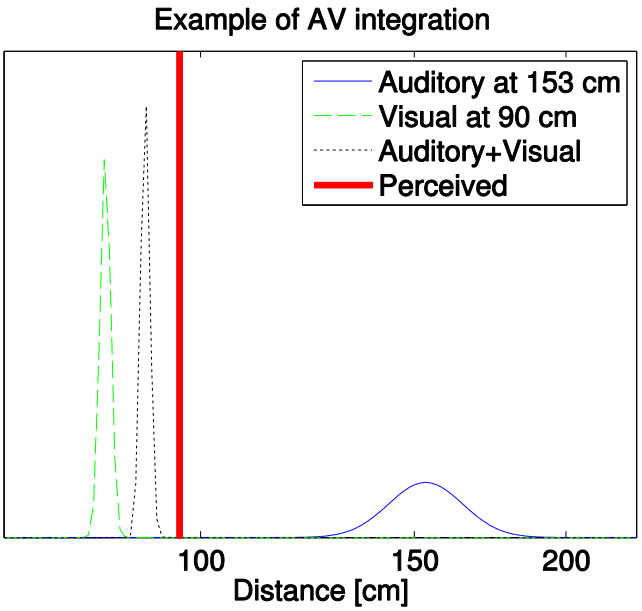

Figure 4-1 Example of ideal observer model (MLE) of auditory and visual stimuli used in the experiment and the actual mean response of the subjects in this condition (red line).

Thus the data of the audio visual experiments (Sec. 3) were fit to the Bayesian model of the AV integration (Bresciani et al. 2006) with the coupling prior represented as the Gaussian ridge on the diagonal, which allowed the coupling of the auditory and visual components. That in result could decrease the weight of the visual component and explain the behavioral data.

4.3 Model

The model was proposed with the aim to account for the differences of the predictions of the MLE model and collected data with the assumption that the AV percept was explained by the brain not necessarily as a unique cause. A common approach to model the causal inference is to view the MLE model as a special case of the Bayesian model with the non-informative prior and extend or modify it (Körding et al. 2007). In fact, the Bayesian statistics provides much more flexibility and the structure of the Bayesian model can be of arbitrary complexity while in the MLE model the structure is limited to independent multivariate Gaussian distributions without the prior. The classical Bayesian model applied to audio-visual integration is defined as:

$$P(s_A, s_V | d_A, d_V) = \frac{P(d_A, d_V | s_A, s_V) P(s_A, s_V)}{P(d_A, d_V)} \quad (4)$$

where $P(s_A, s_V|d_A, d_V)$ is the posterior probability of the event s_A, s_V given the observation d_A, d_V . $P(d_A, d_A|s_A, s_A)$ expresses the likelihood function which describes the statistics related to sensors i.e., how the sensors generate the d_{AV} given the percept s_A, s_V and the prior $P(s_A, s_V)$ expresses the statistics of the scene, i.e., expectations about the scene regardless of the actual observations. The term $P(d_A, d_V)$ in this case only represents the normalizing constant (probability must equal to 1), since it is independent of s_A, s_V . In order to be able to link our model with the MLE model the auditory and visual components were assumed to be independent of the visual and auditory components and the likelihood function was modeled as the product of the auditory component, visual component

$$P(s_A, s_V|d_A, d_A) = P(d_A|s_A)P(d_V|s_V)P(s_A, s_V) \quad (5)$$

Which would equal to the MLE model if the $P(s)$ was non-informative and $P(d_A|s_A)$ and $P(d_V|s_V)$ were Gaussians $N(s_A; \mu_A, \sigma_A)$ and $N(s_V; \mu_V, \sigma_V)$, respectively. The μ and σ characterize the mean and standard deviation of the distribution, 's' is the variable. Therefore the complexity of the model is expressed in the likelihood function and the prior function. For example the independence does not have to be assumed in the likelihood function and the co-variance term of the bivariate Gaussian can express the correlation between the two Gaussians (Oruç et al. 2003), or the Gaussian can be replaced by a different distribution (Rosas and Wichmann 2011). However, in the audio-visual integration paradigms, which are often modeled by the causal inference models, the subject of modeling is only the prior function (Körding et al. 2007).

In the current model the likelihood function was modeled similarly as in the MLE model as the product of the Gaussians. However, one way to allow the 'quasi' causal structure of the model is to change the prior $P(s)$ such that it allows coupling of the visual and auditory components which was used previously to model the integration of the visual and tactile information (Bresciani et al. 2006). In the current model the the prior P(s) has a form of the Gaussian ridge on the diagonal

$$P(s_A, s_V) \propto e^{-\frac{(s_A - s_V)^2}{2\sigma_{coupling}^2}} \quad (6)$$

Where $\sigma_{coupling}$ determines the amount of coupling between the visual and auditory components. If the value $\sigma_{coupling}$ approaches zero, the prior is non-informative and the model is identical to the MLE. Therefore the s_A and s_V are completely unified which

means that they are perceived at the same distance. In theory, the shape of the prior is then infinity on the diagonal and zeros elsewhere. As the $\sigma_{coupling}$ increases the s_A and s_V are less and less unified up to the point when they are independent thus the prior would be equal to the likelihood function. The coupling prior is a simple model of the cross modal interaction because in practice it only increases the variance of one of the components. That can be seen when we express the $P(s_A|d_{AV})$ by marginalizing the posterior with respect to s_V (Körding et al. 2007):

$$\begin{aligned} P(s_A|d_{AV}) &\propto \int P(d_A|s_A)P(d_V|s_V)P(s_A,s_V)ds_V \\ &\propto P(s_A|d_A)\int e^{-\frac{(s_A-s_V)^2}{2\sigma^2_{coupling}}}P(d_V|s_A)ds_V \\ &\propto N(s_A;d_A,\sigma_A)\int N\big(s_V;s_A,\sigma_{coupling}\big)N(s_V;d_V,\sigma_V)ds_V \\ &= N(s_A;d_A,\sigma_A)N\big(s_A;d_V,\sqrt{\sigma_V+\sigma_{coupling}}\big) \end{aligned} \tag{7}$$

and since all the probability functions are expressed by the Gaussian then the result is also the Gaussian with the following properties

$$\begin{aligned} &P(s_A|d_{AV}) \\ &\propto N\left(s_A;\frac{d_A\sigma_A^{-2}+d_V(\sigma_V^2+\sigma^2_{coupling})^{-1}}{\sigma_A^{-2}+(\sigma_V^2+\sigma^2_{coupling})^{-1}},\frac{1}{\sqrt{\sigma_A^{-2}+(\sigma_V^2+\sigma^2_{coupling})^{-1}}}\right) \end{aligned} \tag{8}$$

Which is exactly the MLE model described by **(3)** with the increased variability of the visual component. If we wanted to express the $P(s_V|d_{AV})$ we need to simply interchange A and V. The estimate $\widehat{s_A}$ can be expressed as the mean of the **(8)**.

The model with the coupling prior and the MLE model are visualized on **Figure 4-2** using the same example as on **Figure 4-1**, the auditory target at 153 cm and the visual target at 90 cm. The figure visualizes the likelihood function (A) which is composed of the bivariate Gaussian with co-variance matrix

$$\Sigma_{likelihood} = \begin{bmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_V^2 \end{bmatrix} \tag{9}$$

The function plotted on the graph (A) expresses the perceived distance and standard deviations σ_A and σ_V of the visual and auditory components when presented in

isolation plotted in the logarithmic space. The co-variance between the visual and auditory components was set to 0. The upper two panels (B, C) express the prior function and the posterior estimate of the MLE model of the AV integration. They demonstrate that the optimally combined percept must be on the diagonal of the plot because both auditory and visual components are always perceived at the same distance with the same estimate of the standard deviation σ_{AV}. In theory, the vetoing model (dominance) would be if the resulting distribution was only a shift of the likelihood to the left or up (to the diagonal). The shift to the left would be visual dominance, the shift up would be auditory dominance. The bottom two panels show (D) the Gaussian ridge on diagonal, i.e. the coupling prior, with $\sigma_{coupling}$ denoting the amount of coupling and (E) the posterior estimate of the Bayesian model with the coupling prior. It shows that the resulting percept is not on the diagonal, which means that the auditory and visual components were not perceived at the same distance. Theoretically if the $\sigma_{coupling}$ increased, the position of the posterior estimate would shift towards the position of the likelihood function even more, in infinity of coupling it would reach the exact position of the likelihood function.

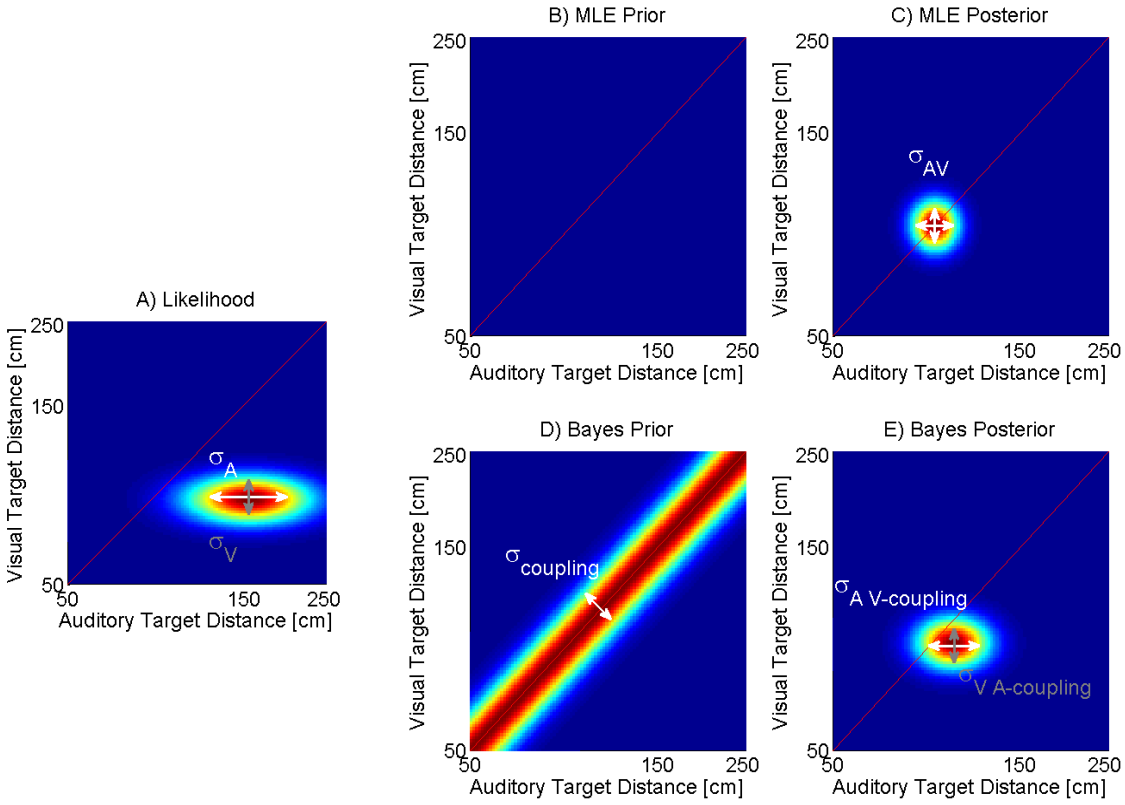

Figure 4-2 Visualization of the MLE and Bayesian model with the coupling prior. The auditory component was perceived at distance of 153 cm and the visual component was perceived on 90 cm. (A) Likelihood function is a bivariate Gaussian

with the variances corresponding to the actual perceptual estimates σ_A a σ_V. (B) Non-informative prior of the MLE model which results in the fusion of the two components always on the diagonal. (C) Posterior estimate of the MLE model. Both components are perceived on the diagonal, i.e., with the same distance and with the equal variance σ_{AV}. (D) Coupling prior is the Gaussian ridge on the diagonal, $\sigma_{coupling}$ expresses the amount of the coupling. (E) The estimate of the Bayesian model with the coupling prior. The A and V components are not perceived at equal distances, the amount of disconcordance and $\sigma_{A\,V-coupling}$ ((8) and $\sigma_{V\,A-coupling}$ is determined by the $\sigma_{coupling}$ and standard deviations of individual components.

The actual MLE model was based on the measurements of the Experiment 3 (Sec. 3.6.1) and Experiment 4 (Sec. 3.6.2) described in the previous chapter, and the subject's performance in Experiment 1 (Sec. 3.4) and Experiment 2 (Sec. 3.5) was compared to the predictions of this model. Experiment 3 and Experiment 4 measured the perception of the auditory and visual stimuli presented in isolation. The mean perceived distance and within-subject standard deviation of both visual and auditory components were obtained from the responses in these unimodal experiments. The across-subject mean perceived distance in both modalities was fit with the power model (Sec. 3.6). These unimodal model fits together with measured standard deviations were used to obtain the estimates of the mean response of the LME model (**(2)**) in all three conditions of the previous experiments: V-Closer, V-Farther, and V-Aligned conditions (together 24 numbers = 3 conditions x 8 targets). These numbers were used to produce the predictions of the ventriloquism effect (by computing V-Misaligned - V-Aligned) and these 24 numbers were compared to the performance of each subject by computing the variance explained by this model. The variance explained was computed for each subject by comparing the actual performance in the AV trials with the model predictions. The performance was obtained by computing the mean perceived distance in each of the conditions. In Experiment 1, for each subject 24 numbers were obtained in 3 conditions such that the V-Misaligned data (16 numbers) were taken from the adaptation runs and V-Aligned data were taken from the across-session average of the runs 11. In Experiment 2, only 16 numbers were obtained for each subject because the adaptation runs involved both V-Aligned and one of the two V-Misaligned conditions. Therefore each subject was compared only to one of V-Misaligned conditions.

Additionally, the data from the audio visual experiments were fit also the Bayesian model with the coupling prior. Data were fit for each subject with the only parameter $\sigma_{coupling}$ in the means of the least squares. The procedure was similar to the MLE model, the same data were used to compute the predictions. However, instead of directly computing the posterior estimates, the 24 or 16 values of each subject were fit to the mean of the normal distribution ($\widehat{s_A}$) described by **Equation (8)** using the bounded version of nonlinear least squares estimate (MATLAB implementation of the 'trusted-region-reflective' algorithm). The bounds were set to 0 and 0.3.

The aim of the modeling was to explain the audio-visual integration in distance dimension. In the behavioral experiment we estimated the efficacy of the AV integration as the difference of the V-Aligned and V-Misaligned conditions (the magnitude of the ventriloquism effect), therefore here the estimates in these conditions were subtracted to be easily compared to the behavioral data on **Figure 3-5**. **Figure 4-3** describes the results of the modeling in a similar format. As was expected the results of the MLE model (thin dashed lines) were strongly influenced by the visual component. The V-Closer MLE was almost dominated by the visual component while the V-Farther MLE was biased to the lesser extent (the visual components are not shown on **Figure 4-3** because it uses the auditory perceptual baseline, however, the position can be inferred form **Figure 3-3**). The difference between the conditions relates to the compression of the visual component and to lesser extent of the compression of the auditory component (that could be seen in Experiment 3 (Sec. 3.6.1) and Experiment 4 (Sec. 3.6.2), respectively. The jump in the middle of the response range relates to the experimental setup and the exact position of the LEDs (note that the LEDs were not perfectly aligned at 30% disparity). The figure also shows that the MLE model could explain 86% of observed variance in Experiment 1 and 62% of the experimental variance in Experiment 2, however, the predictions are substantially biased from the behavioral data towards visual targets. On the other hand, the predictions of the Bayesian model with the coupling prior are much closer to the real measurements which is also reflected in the increased portions of the explained experimental variance (88% in Experiment 1 and 75% in Experiment 2) and the difference between the MLE model and Bayesian model was also confirmed statistically on the dataset pooled across both experiment (paired two-sided t-test: $p<0.05$). The magnitude of the coupling priors were 0.097 in Experiment 1 and 0.103 and Experiment 2, which is the approximate magnitude of the visual component 0.09 (from Experiment 4) thus the

standard deviation of the coupled visual component was 0.19, which is slightly more of the the actually measured magnitude of the response standard deviation in Experiment 1 and Experiment 2 in AV trials (note that the σ_A was not estimated form the Experiment 1 and Experiment 2), while the standard deviation of the auditory component was 0.26 (from Experiment 3). The coupled visual standard deviation of 0.19 and auditory standard deviation 0.26, according to the **Equation (2)**, leads to the perceptual weight of 0.34 of the auditory component while in the MLE model it would be 0.11, therefore the auditory component had to be weighted more than 3 times more than at the case of the optimal integration.

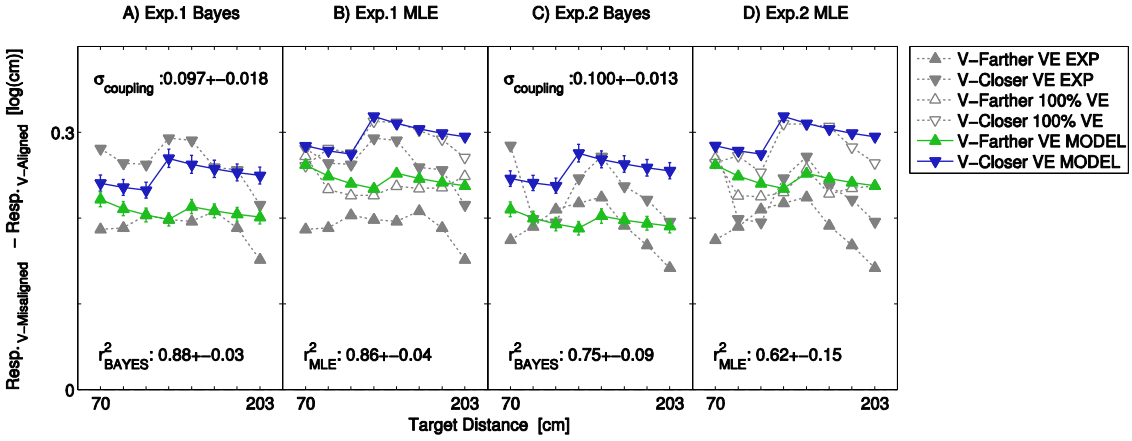

Figure 4-3 Predictions of the AV integration in distance by (A, C) the Bayesian model with the coupling prior and by (B, D) the MLE model, modeled data are shown in color. The MLE model estimates were based on the observations of AV Experiment 3 and AV Experiment 4. The Bayesian model is a modification of the MLE model in which each subject was fitted with one parameter $\sigma_{coupling}$ that represented the width of the Gaussian ridge on diagonal in the prior function. The modeled data in V-Aligned condition were subtracted form V-Misaligned data. The figure also shows the behavioral data (dotted lines with full gray symbols), and the predictions of the complete VE (dotted lines with open gray symbols). The r^2 (±SEM) values express the across-subject mean amount of experimental variance explained by the model. The $\sigma_{coupling}$ (±SEM) values show the across-subject mean estimate of the parameter.

In relation to the actual experimental data, the data of Bayesian model successfully captured the mean bias of the ventriloquism effect, while the mean predictions of the MLE model follow (almost exactly) the model of the 100% VE from the previous chapter (see Sec. 3.4.1.1). The 100% VE model was computed from the behavioral estimates of

the V-Aligned condition therefore it makes sense that the 100% VE corresponds to the MLE model because the MLE expects that the percepts are fully fused. On the other hand the Bayesian model does not fully capture the variation of the VE behavioral data with respect to distance. The behavioral data show much greater peak in the middle of the response range especially in V-Farther in Experiment 2 and V-Closer Experiment 1, however, these discrepancies may result from the across-subject variance and imperfection of the power fit to the auditory data from Experiment 3. Overall, the Bayesian model with the coupling prior explains the observed behavioral data in terms of biases and standard deviations.

4.4 Discussion

The brain combines the auditory and visual spatial information in distance dimension in the similar way as any information from different modalities and dimensions. Our data provide further evidence that the perceptual integration follows the linear weighted combination model, albeit in our experiments the weights of the audio-visual stimuli in distance dimension did not follow the optimal combination rule. It is likely that the simultaneous presentation of the visual and auditory component did not result in the completely fused percept in which both components were perceived as the result of the single cause, rather the auditory and visual components were perceived slightly biased from each other when they were presented with the spatial disparity. The alternative explanation is that the percepts of the visual and auditory components were actually fused but the A component received higher weight.

The current results therefore support the causal inference modeling although the current model does not truly simulate the model of causal inference as it was demonstrated by various other researchers (Körding et al. 2007; Wozny and Shams 2011a). The structure of the their causal inference model involved the hidden variable which explicitly modeled the fact that the percept can be fully integrated or fully segregated (Körding et al. 2007; Wozny and Shams 2011a). In each of the possibilities (integration or segregation), then the percepts are optimally combined according to the LME model. This explicit modeling of the causal structure actually simulates the influence of the disparity on the integration and segregation, i.e., high disparity leads to more probable segregation, lower disparity leads to more probable integration (Körding et al. 2007). In contrast, the current model (Bresciani et al. 2006) did not specifically take into account the disparity of the two components and it only forces the certain segregation

(by introducing the $\sigma_{coupling}$) regardless of the disparity. The reason why the model with the coupling prior was sufficient to explain our data is that the disparity in current experiment was set either to 0% or to 30% of the reference distance and the model was fitting only these two disparities. In the case of the 0% disparity the coupling prior did not negatively influence the prediction because the physical stimuli were at the same location and the $\sigma_{coupling}$ does not change the prediction at this particular condition, therefore the $\sigma_{coupling}$ could model the disparity at 30% disparity. If our experiments systematically manipulated the disparity (Wozny and Shams 2011a; Körding et al. 2007), the coupling prior would be insufficient to capture the effect of segregation at the different levels of disparity and model with the direct causal structure would be more appropriate. It could also be argued that at the 0% of disparity the auditory and visual components were not perfectly aligned on the perceptual space, as was used in the current modeling. From this point of view the direct modeling of the causal structure would perform better.

Taken together a simple Bayesian model outperformed the MLE model in fitting the data of the audio-visual experiments in distance dimension because in comparison the the MLE model it involves a coupling prior which is sufficient to explain the current data because the relative audio-visual disparity was fixed during the experiment. This result can pinpoint to the integration mechanism that brain uses to process the audio visual stimuli presented in distance dimension. The model suggested the interaction between the visual and auditory components although it did not provide the explanation of the biological nature of the interaction. The reason of the observed interaction could originate from the procedural aspects of the experiment because the subjects were explicitly instructed to ignore the visual component and respond only to the auditory component. The subject's responses could be also affected by the response and decision noise because it took some time and relatively complex motor action to respond. Furthermore, the interaction could be truly perceptual although that is less likely since the stimuli were presented from different modalities. Nevertheless, the future experiments should establish how different distance cues in different modalities interact and contribute to the observed phenomenon.

A side note of the current modeling is that the model also tests the assumptions of the constant standard deviations of the auditory and visual components and power model fits that were used by the model (estimated in Experiment 3 and Experiment 4). The model performs well on the logarithmic scale with the constant standard deviations of

both components. For the visual component, the SD was estimated from the fit to the power model. The SD for the auditory component was estimated for each target separately and then averaged afterwards. The averaging of the A SDs introduced some error into the predictions because the behaviorally measured SD varied with distance. Nevertheless, the aim was to minimize the number of parameters to make the model as simple as possible. On the other hand, if the SDs systematically varied with target distance, for instance if the visual standard deviation increased with distance than the model would predict higher weights at the end of the response range, which is not the case. Therefore the model also supports the earlier observations (Kopčo et al. 2012; Anderson and Zahorik 2014; Zahorik et al. 2005) that the visual and auditory distance dimensions are represented on the logarithmic space.

The previous study used the causal inference models to explain the VAE (Wozny and Shams 2011a). They fit their model to the distributions of the pre-adaptation responses and post-adaptation responses and compared the change of parameters of the model. The measures involved the A, V, and AV trials with systematically manipulated amount of disparity. Their model involved seven parameters, which represented the offset and width of distributions of the likelihood function (auditory a visual components) and prior function, which were all represented by the Gaussians (together 6 parameters). The seventh parameter explicitly modeled the probability of fusion of the AV stimuli. Their results showed that the VAE most likely originates from the change of the offset parameters of the auditory component in the likelihood function i.e., that the representation of the auditory space shifted in the direction of the AV training. In our experiments, the pre-adaptation and post-adaptation trials involved only the A trials therefore the VAE can be easily modeled as the change of the mean of the distribution. However, the change itself does not explain the relationship between the VAE and VE. One possible way how to model the VAE is to assume that the VE influences the prior function in the A trials, such that it creates the expectation about the scene that A components are most likely originate from the position which is shifted in direction of the AV disparity. The prior, if modeled by Gaussian, would have 2 parameters (mean and stand deviation). Further, it would be reasonable to assume that the mean of the prior can relate distribution of stimuli in the AV trials and the standard deviation can represent the memory noise which can be modeled by a decay function (if we had enough data in various time intervals). In such way, the VAE can be directly predicted from the VE.

Finally, the Bayesian framework provides high flexibility in modeling multisensory integration due to its stochastic nature. In its core principle it includes a likelihood part that can represent the sensorial processing which it can model changes in the cue extraction phase (as in perceptual learning); however, it also involves the prior function which can model various situations in which the sensory cues are reorganized as a response to learning (cue reweighting). That can be practical in modeling learning of auditory distance cues which was presented in the previous chapter (Sec. 2). This two-step approach to modeling is also consistent with the other previous models of multi-sensory integration, which assumed that perception is influenced by the sensorial noise and memory noise (Shinn-Cunningham 2000a), which in context of Bayesian modeling corresponds likelihood and prior functions.

5 Conclusions

Two experimental studies and one modeling study were conducted in order to investigate the perceptual mechanisms used by the brain spontaneously learns over several days egocentric distance information obtained from reverberation and sound level and to investigate how auditory distance perception is influenced by the visual information.

The primary cues for the auditory distance are sound level and reverberation (Zahorik et al. 2005). Sound level provides relative information about egocentric distance therefore the subject needs to have a priori knowledge about the sound source in order to perceive egocentric distance correctly, whereas reverberation provides absolute information about the distance of the sound source (Mershon and King 1975) and the subject needs to learn only the offset of the particular room (Kopčo et al. 2012) because the auditory distance cues that relate to reverberation vary from room to room. Thus each time we enter a new room the perception must recalibrate.

The first study investigated whether people retain the acoustical memory of the room when they are trained in the auditory distance localization task without feedback over several days in the same room. We therefore wanted to find out whether subjects learn the reverberation related cues of the particular room. The experimental hypothesis was that the subjects would learn the reverberation related cues if they were forced to rely on the intensity independent (relative) cues and conversely they would not learn if the relative cues (such as sound level) was available as the cue for egocentric distance. The hypothesis was not confirmed because the subjects learned to use reverberation cues even when they were trained in the condition with available sound level cue (F). Even more, the learning transferred to the condition in which the sound level cues were not available (R). Subjects learned also when they were trained in the R condition but learning did not transfer to F. The most likely explanation for these findings is that the subjects were actually using the reverberation cues when the sound level cues were available. Nevertheless, there seems to be large variability between the subjects in how they use the sound level cues. Although our experiments demonstrates that people learn auditory distance perception over several days, our analysis revealed that the patterns of improvements were influenced by interleaving the F and R runs within one session. While the F performance improved between the training sessions, the R performance did not improve between training sessions. It only improved between the training sessions and

testing sessions in which the F runs were included. Which means, that F provided a form of calibration for the R condition. That points to the fact that relative distance cues play a crucial role in auditory distance learning. In addition to that, we observed that the R performance improved rapidly during the first testing sessions although the improvement was influenced also by the initial condition (R or F) of the testing session, which suggests that the adaptation process was happing very quickly and perhaps that influenced also long term learning.

An important question for future research is whether people have one representation of the acoustical space or whether each room have its own representation, as well as the question of which acoustical (or other) feature is responsible for learning auditory distance and how absolute and relative cues contribute to learning. It is also not clear whether observed learning generalize across locations within one room, and whether learning also affected the perceptual representation i.e., whether it can improve the perception per se. Taken together, this study provides an evidence for the novel training paradigm (Shinn-Cunningham 2000b; Kopčo et al. 2004b; Schoolmaster et al. 2003, 2004) in the auditory localization task which is consistent with the general condition of the contextual plausibility in the auditory learning (Weinberger 2015).

The second study investigated how vision influences auditory distance perception when the sounds are paired systematically with visual stimuli. The aim of the study was to investigate the asymmetry of ventriloquism effect and aftereffect in distance dimension. Several previous studies investigated the ventriloquism effect (visual capture) in distance dimension, however, this study was the first which investigated systematically ventriloquism aftereffect in distance. The results confirmed the asymmetry between the visual adaptors that were placed in front of the auditory target (V-Closer) and behind the auditory target (V-Farther). The localization bias in the V-Closer AV trials followed the displacement of the visual adaptor, while the V-Farther AV bias decreased dramatically when the distance of auditory target increased. However, the study also showed that the asymmetrical pattern can be accounted to the performance in the perceptual baseline when the sound and light were aligned at distance of auditory targets (V-Aligned). The study also found a systematic displacement in the interleaved A trials which persisted minutes after the AV training, which demonstrated the ventriloquism aftereffect. The aftereffect reached the magnitude of 40-50 % of the ventriloquism effect's magnitude. These results provide insight into processing and integration of the bimodal information

in distance dimension. The knowledge of the cognitive processes in perception of distance has important implications, for example for the virtual reality systems. The auditory and visual distance cues should provide consistent spatial information in 3D reality systems to create real illusion.

The data of the audio-visual study were modeled by the optimal MLE model (Alais and Burr 2004) and by the Bayesian model with the coupling prior (Bresciani et al. 2006). The Bayesian model was successful in explaining the behavioral data. The Bayesian model is a modification of the MLE model in which the coupling term, expresses the degree of integration, increases variability of one component relative to the other. While in the MLE model the fusion is always complete, the model with the coupling prior simulates the decrease of fusion of the auditory and visual component. This means that the model enables the situation in which the brain explained the audio-visual event as two independent events. Although this particular model does not fully imitate the causal inference (Körding et al. 2007), it simulates a form of the causal inference model which only decreases the amount of integration and sets it to the fixed value. On the other hand, the causal inference models can take into account the actual disparity of the two components and increase the influence of integration as a function of actual disparity. However, in the current experiments the relative disparity was fixed, therefore there was no need to capture the effect of increasing disparity. Taken together, the mathematical modeling suggests that the visual and auditory components in the AV experiments in distance dimension were not completely fused, which resulted in lower perceptual weights of the visual component and higher perceptual weights of the auditory component. Nevertheless, the behavioral results of the AV integration can be explained by the perceptual properties of the underlying cues when the model allows that the two events can interact.

Finally, the experiments and a modeling showed that studying auditory perception can have a broader impact on understanding the perceptual and cognitive processes in the human brain. The first study showed that people retain the acoustical memories and it pointed to the complex mechanism of the perception of relative and absolute cues in auditory distance perception. The second study showed that the integration of auditory and visual information in distance seems to be asymmetric; however, it can relate to increase of localization blur with distance. The study in modeling showed that the all outcomes of the audio-visual experiments can be understood in terms of the properties of

the underlying cues and that the auditory and visual cues are in interaction. Nevertheless, auditory distance perception is influenced by the experience with room reverberation and the cues from the visual modality. Therefore it is likely the brain incorporates the cues in order to maintain the perceptual stability.

6 Resumé

Dve experimentálne štúdie a matematické modelovanie, ktoré boli predstavené v tejto práci skúmali kognitívne a perceptuálne procesy, využívané mozgom pri spracovaní priestorových informácií o vzdialenosti, ktoré prichádzajú zo sluchovej a zrakovej modality.

Prvá štúdia sa zamerala na spontánny proces učenia vnímania sluchovej vzdialenosti v reverberantnom prostredí. Experimentálne subjekty boli trénované počas viacerých dní v zvukovej lokalizačnej úlohe bez spätnej väzby. Subjekty sa počas experimentu zlepšili, čo naznačuje, že tréning dokázal vylepšiť vnímanie reverberácie, ktorá je dôležitá pre vnímanie sluchovej vzdialenosti. Napriek našim očakávaniam sa ľudia zlepšili i v tom prípade, keď mohli odpovedať podľa intenzity zvuku, ktorá zvyčajne podáva iba relatívnu informáciu o sluchovej vzdialenosti, takže experimentálne subjekty pravdepodobne používali aj tie informácie o polohe zvuku, ktoré súvisia s reverberáciou miestnosti a zároveň aj informácie o intenzite zvuku, čo im pomohlo zlepšiť ich vnímanie vzdialenosti zvukov.

Druhá štúdia skúmala vplyv vizuálnych stimulov na vnímanie sluchovej vzdialenosti. Výsledky ukázali, že bližšie vizuálne adaptory priťahovali vnem sluchovej vzdialenosti viac ako vzdialenejšie vizuálne adaptory, vzhľadom na polohu sluchového cieľa. Tieto výsledky sa čiastočne dajú vysvetliť tým, že vnímanie subjektov boli značne kompresované i v tom prípade, keď sluchové ciele boli prezentované v rovnakej vzdialenosti ako vizuálne adaptory.

Dáta druhej štúdie boli modelované lineárnym vážením sluchovej a vizuálnej informácie. Váha bol vypočítaná buď ako optimálny pomer štandardných odchýlok jednotlivých vnemov alebo bola vypočítaná pomocou podobného modelu, ktorý bol založený na Baysovskej štatistike a v ktorom bol modelovaný možný pokles audio-vizuálnej integrácie. V optimálnom modeli vždy dochádza k úplnej integrácií oboch vnemov, ale experimentálne dáta naznačujú, že tento model ich nemôže celkom vysvetliť. Pomocou Bayesovského modelu sme vedeli vysvetliť 75% a 88% experimentálnej variancie čo bolo štatisticky viac ako sme vedeli vysvetliť pomocou optimálneho modelu. Tieto výsledky naznačujú, že mozog sa snaží odhadnúť kauzálnu štruktúru kombinovaného vnemu, čo môže vysvetliť naše experimentálne pozorovania.

7 Bibliography

ABEL, Sharon M and J.E.S PAIK, 2004. The benefit of practice for sound localization without sight. *Applied Acoustics* [online]. 2004, vol. 65, no. 3, pp. 229–241 [accessed. 14. January 2014]. ISSN 0003682X. Available from: doi:10.1016/j.apacoust.2003.10.003

AHISSAR, Merav and Shaul HOCHSTEIN, 2004. The reverse hierarchy theory of visual perceptual learning. *Trends in cognitive sciences* [online]. 2004, vol. 8, no. 10, pp. 457–64 [accessed. 15. July 2011]. ISSN 1364-6613. Available from: doi:10.1016/j.tics.2004.08.011

AHVENINEN, Jyrki, Norbert KOPČO and Iiro P JÄÄSKELÄINEN, 2014. Psychophysics and neuronal bases of sound localization in humans. *Hearing research* [online]. 2014, vol. 307, pp. 86–97 [accessed. 19. December 2013]. ISSN 1878-5891. Available from: doi:10.1016/j.heares.2013.07.008

ALAIS, David and David BURR, 2004. The ventriloquist effect results from near-optimal bimodal integration. *Current biology : CB* [online]. 2004, vol. 14, no. 3, pp. 257–62 [accessed. 6. August 2013]. ISSN 0960-9822. Available from: doi:10.1016/j.cub.2004.01.029

ALTMANN, Christian F., Kentaro ONO, Akiko CALLAN, Masao MATSUHASHI, Tatsuya MIMA and Hidenao FUKUYAMA, 2013. Environmental reverberation affects processing of sound intensity in right temporal cortex. *European Journal of Neuroscience* [online]. 2013, vol. 38, no. February, pp. 3210–3220. ISSN 0953816X. Available from: doi:10.1111/ejn.12318

ANDERSON, Paul W. and Pavel ZAHORIK, 2014. Auditory/visual distance estimation: accuracy and variability. *Frontiers in Psychology* [online]. 2014, vol. 5, no. October, pp. 1–11. ISSN 1664-1078. Available from: doi:10.3389/fpsyg.2014.01097

ASHMEAD, D H, D LEROY and R D ODOM, 1990. Perception of the relative distances of nearby sound sources. *Perception & psychophysics* [online]. 1990, vol. 47, no. 4, pp. 326–31. ISSN 0031-5117. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2345684

ASHMEAD, D.H., D.L. DAVIS and A. NORTHINGTON, 1995. Contribution of listeners' approaching motion to auditory distance perception. *Journal of Experimental Psychology: Human Perception and Performance*. 1995, vol. 21, no. 2, pp. 239–256.

BACH, Dominik R., John G. NEUHOFF, Walter PERRIG and Erich SEIFRITZ, 2009. Looming sounds as warning signals: The function of motion cues. *International Journal of Psychophysiology* [online]. 2009, vol. 74, no. 1, pp. 28–33. ISSN 01678760. Available from: doi:10.1016/j.ijpsycho.2009.06.004

BATTAGLIA, Peter W P.W., R.A. Robert a JACOBS and Richard N R.N. ASLIN, 2003. Bayesian integration of visual and auditory signals for spatial localization. *Journal of the*

Optical Society of America. A, Optics, image science, and vision [online]. 2003, vol. 20, no. 7, pp. 1391–7. ISSN 1084-7529. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12868643

BERTELSON, P., I. FRISSEN, J. VROOMEN and B. DE GELDER, 2006. The aftereffects of ventriloquism: Patterns of spatial generalization. *Perception and Psychophysics*. 2006, vol. 68, no. 3, pp. 428–436.

BERTELSON, P., J. VROOMEN, B. DE GELDER, J. DRIVER, P. BERTELSON, B. DE GELDER and J. DRIVER, 2000. The ventriloquist effect does not depend on the direction of deliberate visual attention. In: *Perception and Psychophysics* [online]. p. 321–32. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10723211

BLAUERT, J., 1997. *Spatial Hearing*. Cambridge, MA: MIT Press.

BRAASCH, Jonas, 2013. A precedence effect model to simulate localization dominance using an adaptive, stimulus parameter-based inhibition process. *The Journal of the Acoustical Society of America* [online]. 2013, vol. 134, no. 1, pp. 420–35 [accessed. 18. December 2013]. ISSN 1520-8524. Available from: doi:10.1121/1.4807829

BRANDEWIE, Eugene and Pavel ZAHORIK, 2010. Prior listening in rooms improves speech intelligibility. *The Journal of the Acoustical Society of America* [online]. 2010, vol. 128, no. 1, pp. 291–9 [accessed. 15. August 2011]. ISSN 1520-8524. Available from: doi:10.1121/1.3436565

BRESCIANI, Jean-Pierre, Franziska DAMMEIER and Marc O ERNST, 2006. Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision* [online]. 2006, vol. 6, no. 5, pp. 554–64. ISSN 1534-7362. Available from: doi:10.1167/6.5.2

BRIMIJOIN, W. Owen, Alan W. BOYD and Michael a. AKEROYD, 2013. The contribution of head movement to the externalization and internalization of sounds. *PLoS ONE* [online]. 2013, vol. 8, no. 12, pp. 1–12. ISSN 19326203. Available from: doi:10.1371/journal.pone.0083068

BRONKHORST, A.W. and T. HOUTGAST, 1999. Auditory distance perception in rooms. *Nature*. 1999, vol. 397, no. 11 February, pp. 517–520.

BRONKHORST, Adelbert W., 2002. Modeling auditory distance perception in rooms. In: *Proc. EAA ForumAcusticum*.

BROWN, Andrew D, Marina S KUZNETSOVA, William J SPAIN and G Christopher STECKER, 2012. Frequency-specific, location-nonspecific adaptation of interaural time difference sensitivity. *Hearing research* [online]. 2012, vol. 291, no. 1-2, pp. 52–6 [accessed. 31. July 2014]. ISSN 1878-5891. Available from: doi:10.1016/j.heares.2012.06.002

BROWN, Andrew D, G. Christopher STECKER and Daniel J TOLLIN, 2015. The Precedence Effect In Sound Localization. *Journal of the Association for Research in Otolaryngology* [online]. 2015, vol. 16, no. 1, pp. 1–28. Available from: doi:10.1007/s10162-014-0496-2

BROWN, C., 2002. *T60 Matlab function2004* [online]. Available from: http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=1212&objectType=file#

BRUNGART, D.S. and N.I. DURLACH, 1999. Auditory localization of nearby sources II: Localization of a broadband source in the near field. *Journal of the Acoustical Society of America*. 1999, vol. 106, no. 4, pp. 1956–1968.

BRUNGART, D.S. and W.M. RABINOWITZ, 1999. Auditory localization of nearby sources I: Head-related transfer functions. *Journal of the Acoustical Society of America*. 1999, vol. 106, no. 3, pp. 1465–1479.

BRUNGART, Douglas S and Brian D SIMPSON, 2002. The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. *The Journal of the Acoustical Society of America* [online]. 2002, vol. 112, no. 2, pp. 664–676. ISSN 00014966. Available from: doi:10.1121/1.1490592

BRUNS, Patrick, Ronja LIEBNAU and Brigitte RÖDER, 2011. Cross-modal training induces changes in spatial representations early in the auditory processing pathway. *Psychological science* [online]. 2011, vol. 22, no. 9, pp. 1120–6 [accessed. 29. September 2014]. ISSN 1467-9280. Available from: doi:10.1177/0956797611416254

CALCAGNO, Esteban R, Ezequiel L ABREGÚ, Manuel C EGUÍA and Ramiro VERGARA, 2012. The role of vision in auditory distance perception. *Perception* [online]. 2012, vol. 41, no. 2, pp. 175–192 [accessed. 18. December 2013]. ISSN 0301-0066. Available from: doi:10.1068/p7153

CARLILE, Simon, 2014. The plastic ear and perceptual relearning in auditory spatial perception. *Frontiers in Neuroscience* [online]. 2014, vol. 8, no. 8 JUL, pp. 1–13. ISSN 1662453X. Available from: doi:10.3389/fnins.2014.00237

CARLILE, Simon, Stephanie HYAMS and Skye DELANEY, 2001. Systematic distortions of auditory space perception following prolonged exposure to broadband noise. *The Journal of the Acoustical Society of America* [online]. 2001, vol. 110, no. 1, p. 416 [accessed. 18. December 2013]. ISSN 00014966. Available from: doi:10.1121/1.1375843

CATIC, Jasmina, Sébastien SANTURETTE and Torsten DAU, 2015. The role of reverberation-related binaural cues in the externalization of speech. *The Journal of the Acoustical Society of America* [online]. 2015, vol. 138, no. 2, pp. 1154–1167. ISSN 0001-4966. Available from: doi:10.1121/1.4928132

CATIC, Jasmina, Sébastien Sébastien SANTURETTE, Jörg M. Jörg M BUCHHOLZ,

Fredrik GRAN and Torsten DAU, 2013. The effect of interaural-level-difference fluctuations on the externalization of sound. *The Journal of the Acoustical Society of America* [online]. 2013, vol. 134, no. 2, p. 1232. ISSN 00014966. Available from: doi:10.1121/1.4812264

CLIFTON, R K and R L FREYMAN, 1997. The precedence effect: Beyond echo suppression. In: R. GILKEY and T. ANDERSON, eds. *Binaural and spatial hearing in real and virtual environments*. Mahwah, NJ: Lawrence Erlbaum Associates, p. 334–362.

COCHRAN, W. G., 1937. Problems Arising in the Analysis of a Series of Similar Experiments. *Supplement to the Journal of the Royal Statistical Society* [online]. 1937, vol. 4, no. 1, p. 102. ISSN 14666162. Available from: doi:10.2307/2984123

COLEMAN, P.D., 1962. Failure to localize the source distance of an unfamiliar sound. *Journal of the Acoustical Society of America*. 1962, vol. 34, no. 1938, pp. 345–346.

COLEMAN, P.D., 1968. Dual role of frequency spectrum in determination of auditory distance. *Journal of the Acoustical Society of America*. 1968, vol. 44, no. 2, pp. 631–632.

DAHMEN, Johannes C, Peter KEATING, Fernando R NODAL, Andreas L SCHULZ and Andrew J KING, 2010. Adaptation to stimulus statistics in the perception and neural representation of auditory space. *Neuron* [online]. 2010, vol. 66, no. 6, pp. 937–48 [accessed. 13. August 2013]. ISSN 1097-4199. Available from: doi:10.1016/j.neuron.2010.05.018

DEVORE, Sasha, Antje IHLEFELD, Kenneth HANCOCK, Barbara SHINN-CUNNINGHAM and Bertrand DELGUTTE, 2009. Accurate Sound Localization in Reverberant Environments Is Mediated by Robust Encoding of Spatial Cues in the Auditory Midbrain. *Neuron* [online]. 2009, vol. 62, no. 1, pp. 123–134 [accessed. 5. July 2011]. ISSN 08966273. Available from: doi:10.1016/j.neuron.2009.02.018

DEVORE, Sasha, Andrew SCHWARTZ and Bertrand DELGUTTE, 2010. Effect of Reverberation on Directional Sensitivity of Auditory Neurons: Central and Peripheral Factors. In: Enrique A. LOPEZ-POVEDA, Alan R. PALMER and Ray MEDDIS, eds. *The Neurophysiological Bases of Auditory Perception* [online]. New York, NY: Springer New York, p. 273–282. ISBN 978-1-4419-5685-9. Available from: doi:10.1007/978-1-4419-5686-6_26

DIETZ, Mathias, Stephan D. EWERT and Volker HOHMANN, 2011. Auditory model based direction estimation of concurrent speakers from binaural signals. *Speech Communication* [online]. 2011, vol. 53, no. 5, pp. 592–605 [accessed. 24. November 2011]. ISSN 01676393. Available from: doi:10.1016/j.specom.2010.05.006

DRIVER, J. and C. SPENCE, 1998. Crossmodal attention. *Curr Opin Neurobiol* [online]. 1998, vol. 8, no. 2, pp. 245–53. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9635209

ERNST, M.O. and M.S. BANKS, 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* [online]. 2002, vol. 415, no. 6870, pp. 429–33. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11807554

EŠTOČINOVÁ, Jana, Jyrki AHVENINEN, Samantha HUANG, Stephanie ROSSI and Norbert KOPČO, 2015. Auditory Distance Perception with Congruent and Incongruent Cues. In: *38th MidWinter meeting of the Association for Research in Otolaryngology*.

FALLER, Christof and Juha MERIMAA, 2004. Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *The Journal of the Acoustical Society of America* [online]. 2004, vol. 116, no. 5, p. 3075 [accessed. 14. June 2011]. ISSN 00014966. Available from: doi:10.1121/1.1791872

FLUITT, Kim F, Timothy MERMAGEN and Tomasz LETOWSKI, 2013. Auditory Perception in Open Field : Distance Estimation. 2013, no. July.

GARDNER, M.B., 1968. Proximity image effect in sound localization. *Journal of the Acoustical Society of America*. 1968, vol. 43, no. 6, p. 163.

GEORGANTI, E., T. MAY, S. VAN DE PAR and J. MOURJOPOULOS, 2013. Extracting Sound-Source-Distance Information from Binaural Signals. In: *The Technology of Binaural Listening* [online]. Berlin, Heidelberg, Heidelberg: Springer Berlin Heidelberg, p. 171–199. Available from: doi:10.1007/978-3-642-37762-4_7

GRAZIANO, M.S. S, L.A.J. a REISS and C.G. G GROSS, 1999. A neuronal representation of the location of nearby sounds. *Nature* [online]. 1999, vol. 397, no. 6718, pp. 428–430. ISSN 0028-0836. Available from: doi:10.1038/17115

GROTHE, Benedikt, Michael PECKA and David MCALPINE, 2010. Mechanisms of Sound Localization in Mammals. *Physiological Reviews* [online]. 2010, vol. 90, no. 3, pp. 983–1012. ISSN 0031-9333. Available from: doi:10.1152/physrev.00026.2009

HALL, Deborah a. and David R. MOORE, 2003. Auditory neuroscience: The salience of looming sounds. *Current Biology* [online]. 2003, vol. 13, no. 3, pp. 91–93. ISSN 09609822. Available from: doi:10.1016/S0960-9822(03)00034-4

HARTMANN, W. M., 1983. Localization of sound in rooms. *The Journal of the Acoustical Society of America* [online]. 1983, vol. 74, no. 5, p. 1380. ISSN 00014966. Available from: doi:10.1121/1.390163

HARTMANN, William Morris, 1989. Localization of sound in rooms IV: The Franssen effect. *The Journal of the Acoustical Society of America* [online]. 1989, vol. 86, no. 4, p. 1366 [accessed. 17. May 2012]. ISSN 00014966. Available from: doi:10.1121/1.398696

HERRON, Timothy, 2005. *C Language Exploratory Analysis of Variance with Enhancements* [online]. 2005. B.m.: HCN Laboratory, UC Davis & VANCHCS,

Martinez, CA, USA. Available from: http://www.ebire.org/hcnlab/software/cleave.html

HOFMAN, P.M., J.G.A. VAN RISWICK and A.J. VAN OPSTAL, 1998. Relearning sound localization with new ears. *Nature Neuroscience*. 1998, vol. 1, no. 5, pp. 417–421.

HUNG, Shao-Chin and Aaron R SEITZ, 2014. Prolonged training at threshold promotes robust retinotopic specificity in perceptual learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* [online]. 2014, vol. 34, no. 25, pp. 8423–31. ISSN 1529-2401. Available from: doi:10.1523/JNEUROSCI.0745-14.2014

CHAN, Chetwyn C.H., Alex W.K. WONG, Kin-Hung TING, Susan WHITFIELD-GABRIELI, Jufang HE and Tatia M.C. LEE, 2012a. Cross auditory-spatial learning in early-blind individuals. *Human Brain Mapping* [online]. 2012, vol. 33, no. 11, pp. 2714–2727. ISSN 10659471. Available from: doi:10.1002/hbm.21395

CHAN, Jason S, Corrina MAGUINNESS, Danuta LISIECKA, Annalisa SETTI and Fiona N NEWELL, 2012b. Evidence for crossmodal interactions across depth on target localisation performance in a spatial array. *Perception* [online]. 2012, vol. 41, no. 7, pp. 757–773. ISSN 0301-0066. Available from: doi:10.1068/p7230

IHLEFELD, Antje and Barbara G. SHINN-CUNNINGHAM, 2011. Effect of source spectrum on sound localization in an everyday reverberant room. *The Journal of the Acoustical Society of America* [online]. 2011, vol. 130, no. 1, p. 324 [accessed. 20. July 2011]. ISSN 00014966. Available from: doi:10.1121/1.3596476

JACK, Charles E. and Willard R. THURLOW, 1973. Effects of degree of visual association and angle of displacement on the“ ventriloquism” effect. *Perceptual and Motor Skills* [online]. 1973, vol. 37, no. 3, pp. 967–979. ISSN 0031-5125. Available from: doi:10.2466/pms.1973.37.3.967

KACELNIK, Oliver, Fernando R NODAL, Carl H PARSONS and Andrew J KING, 2006. Training-induced plasticity of auditory localization in adult mammals. *PLoS biology* [online]. 2006, vol. 4, no. 4, p. e71 [accessed. 17. July 2011]. ISSN 1545-7885. Available from: doi:10.1371/journal.pbio.0040071

KASHINO, Makio and Shin’ya NISHIDA, 1998. Adaptation in the processing of interaural time differences revealed by the auditory localization aftereffect. *The Journal of the Acoustical Society of America* [online]. 1998, vol. 103, no. 6, p. 3597 [accessed. 18. December 2013]. ISSN 00014966. Available from: doi:10.1121/1.423064

KEEN, Rachel and Richard L FREYMAN, 2009. Release and re-buildup of listeners’ models of auditory space. *The Journal of the Acoustical Society of America* [online]. 2009, vol. 125, no. 5, pp. 3243–52 [accessed. 13. August 2013]. ISSN 1520-8524. Available from: doi:10.1121/1.3097472

KIM, Duck O, Pavel ZAHORIK, X Laurel H CARNEY, Brian B BISHOP and Shigeyuki KUWADA, 2015. Auditory Distance Coding in Rabbit Midbrain Neurons and Human Perception : Monaural Amplitude Modulation Depth as a Cue [online]. 2015, vol. 35, no.

13, pp. 5360–5372. Available from: doi:10.1523/JNEUROSCI.3798-14.2015

KING, Andrew J, Johannes C DAHMEN, Peter KEATING, Nicholas D LEACH, Fernando R NODAL and Victoria M BAJO, 2011. Neural circuits underlying adaptation and learning in the perception of auditory space. *Neuroscience and biobehavioral reviews* [online]. 2011, vol. 35, no. 10, pp. 2129–39 [accessed. 18. September 2014]. ISSN 1873-7528. Available from: doi:10.1016/j.neubiorev.2011.03.008

KNILL, David C., 2007. Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of VIsion* [online]. 2007, vol. 1, pp. 1–24. Available from: doi:10.1167/0.0.1

KNUDSEN, E.I., 2002. Instructed learning in the auditory localization pathway of the barn owl. *Nature*. 2002, vol. 417, pp. 322–328.

KOLARIK, Andrew J, Silvia CIRSTEA and Shahina PARDHAN, 2013a. Evidence for enhanced discrimination of virtual auditory distance among blind listeners using level and direct-to-reverberant cues. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale* [online]. 2013, vol. 224, no. 4, pp. 623–33 [accessed. 13. August 2013]. ISSN 1432-1106. Available from: doi:10.1007/s00221-012-3340-0

KOLARIK, Andrew J., Silvia CIRSTEA, Shahina PARDHAN and Brian C. MOORE, 2013b. An assessment of virtual auditory distance judgments among blind and sighted listeners [online]. 2013, vol. 19, pp. 050043–050043 [accessed. 13. December 2013]. Available from: doi:10.1121/1.4799570

KOLARIK, Andrew J., Brian C. J. MOORE, Pavel ZAHORIK, Silvia CIRSTEA and Shahina PARDHAN, 2015. Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, & Psychophysics* [online]. 2015. ISSN 1943-3921. Available from: doi:10.3758/s13414-015-1015-1

KOPČO, N, D ČELJUSKA, Miroslav PUSZTA, N. KOPCO, D. CELJUSKA, Miroslav PUSZTA, M. RACEK and M. SARNOVSKY, 2004a. Effect of spectral content and learning on auditory distance perception. *Proc. 2nd Slovak-Hungarian …* [online]. 2004, pp. 1–7. Available from: http://neuron-ai.tuke.sk/~celjuska/papers/kopcosami.pdf

KOPČO, Norbert, Virginia BEST and Barbara G. SHINN-CUNNINGHAM, 2007. Sound localization with a preceding distractor. *The Journal of the Acoustical Society of America* [online]. 2007, vol. 121, no. 1, p. 420 [accessed. 22. March 2014]. ISSN 00014966. Available from: doi:10.1121/1.2390677

KOPČO, Norbert, Samantha HUANG, John W. BELLIVEAU, Tommi RAIJ, Chinmayi TENGSHE, Jyrki AHVENINEN, N. KOPCO, Samantha HUANG, John W. BELLIVEAU, Tommi RAIJ, Chinmayi TENGSHE and Jyrki AHVENINEN, 2012. Neuronal representations of distance in human auditory cortex. *Proceedings of the National Academy of Sciences* [online]. 2012, vol. 109, no. 27, pp. 11019–11024

[accessed. 31. October 2012]. ISSN 0027-8424. Available from: doi:10.1073/pnas.1119496109

KOPČO, Norbert, I-Fan LIN, Barbara G SHINN-CUNNINGHAM and Jennifer M GROH, 2009. Reference frame of the ventriloquism aftereffect. *The Journal of neuroscience : the official journal of the Society for Neuroscience* [online]. 2009, vol. 29, no. 44, pp. 13809–14 [accessed. 22. August 2011]. ISSN 1529-2401. Available from: doi:10.1523/JNEUROSCI.2783-09.2009

KOPČO, Norbert, Ľuboš MARCINEK, Beáta TOMORIOVÁ and Ľuboš HLÁDEK, 2015. Contextual plasticity, top-down, and non-auditory factors in sound localization with a distractora). *The Journal of the Acoustical Society of America* [online]. 2015, vol. 137, no. 4, pp. EL281–EL287. ISSN 0001-4966. Available from: doi:10.1121/1.4914999

KOPČO, Norbert and Barbara G. SHINN-CUNNINGHAM, 2011. Effect of stimulus spectrum on distance perception for nearby sources. *The Journal of the Acoustical Society of America* [online]. 2011, vol. 130, no. 3, p. 1530 [accessed. 24. November 2011]. ISSN 00014966. Available from: doi:10.1121/1.3613705

KOPČO, Norbert, Pierre SILVERA, Beáta TOMORIOVÁ, Aaron SEITZ, M. SCHOOLMASTER and B.G. SHINN-CUNNINGHAM, 2004b. Learning to judge distance of nearby sounds in reverberant and anechoic environments [online]. 2004, p. 14 [accessed. 5. January 2012]. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.72.2952&rep=rep1&type=pdf

KOPČO, Norbert, Eleni VLAHOU, Ueno KANAKO and Barbara SHINN-CUNNINGHAM, 2013. Exposure to Consistent Room Reverberation Facilitates Consonant Perception. 2013, p. 835976.

KÖRDING, Konrad P, Ulrik R BEIERHOLM, Wei Ji MA, Steven R QUARTZ, Joshua B TENENBAUM and Ladan SHAMS, 2007. Causal inference in multisensory perception. *PLoS One* [online]. 2007, vol. 2, no. 9, p. e943. ISSN 1932-6203. Available from: doi:10.1371/journal.pone.0000943

KUMPIK, Daniel P, Oliver KACELNIK and Andrew J KING, 2010. Adaptive reweighting of auditory localization cues in response to chronic unilateral earplugging in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience* [online]. 2010, vol. 30, no. 14, pp. 4883–4894. ISSN 0270-6474. Available from: doi:10.1523/JNEUROSCI.5488-09.2010

KUWADA, Shigeyuki, Duck O. KIM, Kelly-Jo KOCH, Kristina S. ABRAMS, Fabio IDROBO, Pavel ZAHORIK and Laurel H. CARNEY, 2015. Near-Field Discrimination of Sound Source Distance in the Rabbit. *Journal of the Association for Research in Otolaryngology* [online]. 2015, vol. 262, pp. 255–262. ISSN 1525-3961. Available from: doi:10.1007/s10162-014-0505-5

LANDY, M S, L T MALONEY, E B JOHNSTON and M YOUNG, 1995. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision research* [online]. 1995, vol. 35, no. 3, pp. 389–412. ISSN 0042-6989. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7892735

LANDY, Michael S., Martin S. BANKS and David C. KNILL, 2011. Ideal-Observer Models of Cue Integration. In: *Sensory Cue Integration* [online]. B.m.: Oxford University Press, p. 5–29. ISBN 9780199600458. Available from: doi:10.1093/acprof:oso/9780195387247.003.0001

LARSEN, Erik, Nandini IYER, Charissa R LANSING and Albert S FENG, 2008. On the minimum audible difference in direct-to-reverberant energy ratio. *The Journal of the Acoustical Society of America* [online]. 2008, vol. 124, no. 1, pp. 450–61 [accessed. 24. January 2014]. ISSN 1520-8524. Available from: doi:10.1121/1.2936368

LECHNER, H.A., L.R. SQUIRE and J.H. BYRNE, 1999. 100 years of consolidation - remembering Müller and Pilzecker. *Learning & Memory*. 1999, vol. 6, pp. 77–87.

LEWALD, Jörg, 2002. Rapid Adaptation to Auditory-Visual Spatial Disparity. *Learning & Memory* [online]. 2002, vol. 9, no. 5, pp. 268–278. ISSN 1072-0502. Available from: doi:10.1101/lm.51402

LITOVSKY, R.Y., H.S. COLBURN, W.A. YOST and S.J. GUZMAN, 1999. The precedence effect. *Journal of the Acoustical Society of America*. 1999, vol. 106, no. 4, pp. 1633–1654.

LITTLE, A D, D H MERSHON and P H COX, 1992. Spectral content as a cue to perceived auditory distance. *Perception* [online]. 1992, vol. 21, no. 3, pp. 405–16 [accessed. 9. January 2014]. ISSN 0301-0066. Available from: http://www.perceptionjournal.com/perception/fulltext/p21/p210405.pdf

LOOMIS, J M, R L KLATZKY, J W PHILBECK and R G GOLLEDGE, 1998. Assessing auditory distance perception using perceptually directed action. *Perception & psychophysics* [online]. 1998, vol. 60, no. 6, pp. 966–80. ISSN 0031-5117. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9718956

LU, Y and M COOKE, 2010. Binaural Estimation of Sound Source Distance via the Direct-to-Reverberant Energy Ratio for Static and Moving Sources. *IEEE Transactions on Audio, Speech, and Language Processing* [online]. 2010, vol. 18, no. 7, pp. 1793–1805. ISSN 1558-7916. Available from: doi:10.1109/TASL.2010.2050687

MACPHERSON, E.A. and J.C. MIDDLEBROOKS, 2002. Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *J Acoust Soc Am* [online]. 2002, vol. 111, no. 5 Pt 1, pp. 2219–36. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12051442

MARR, D. and T. POGGIO, 1977. From understanding computation to understanding

neural circuitry. *Neuroscience Res . Prog. Bull.* 1977, vol. 15, pp. 470–488.

MERSHON, D H, W L BALLENGER, A D LITTLE, P L MCMURTRY and J L BUCHANAN, 1989. Effects of room reflectance and background noise on perceived auditory distance. *Perception* [online]. 1989, vol. 18, no. 3, pp. 403–16 [accessed. 9. January 2014]. ISSN 0301-0066. Available from: http://www.perceptionweb.com/perception/fulltext/p18/p180403.pdf

MERSHON, D H and L E KING, 1975. Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics* [online]. 1975, vol. 18, no. 6, pp. 409–415 [accessed. 9. January 2014]. ISSN 0031-5117. Available from: doi:10.3758/BF03204113

MERSHON, D.H. and J.N. BOWERS, 1979. Absolute and relative cuse for auditory perception of egocentric distance. *Perception*. 1979, vol. 8, pp. 311–322.

MERSHON, D.H., D.H. DESAULNIERS and J. AMERSON, 1980. Visual capture in auditory distance perception: Proximity image effect reconsidered. *Journal of Auditory Research*. 1980, vol. 20, pp. 129–136.

MIDDLEBROOKS, J.C. C and D.M. M GREEN, 1991. Sound localization by human listeners. *Annual review of psychology* [online]. 1991, vol. 42, no. 1, pp. 135–59 [accessed. 16. April 2012]. ISSN 0066-4308. Available from: doi:10.1146/annurev.ps.42.020191.001031

MILLER, J. A., 1947. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *Journal of Acoustic Society of America*. 1947, no. 19, pp. 609–619.

MILLS, A. W., 1958. On the Minimum Audible Angle. *The Journal of the Acoustical Society of America* [online]. 1958, vol. 30, no. 4, p. 237 [accessed. 18. July 2014]. ISSN 00014966. Available from: doi:10.1121/1.1909553

MIN, YK and DH MERSHON, 2005. An Adjacency effect in auditory distance perception. *Acta acustica united with acustica* [online]. 2005, vol. 91, pp. 480–489 [accessed. 10. January 2013]. Available from: http://www.ingentaconnect.com/content/dav/aaua/2005/00000091/00000003/art00010

MOORE, Brian C. J., 2012. *An Introduction to the Psychology of Hearing (6e)*. B.m.: Emerald Group Publishing Limited. ISBN 978-1-78052-038-4.

MOORE, David R and Andrew J KING, 1999. Auditory perception: The near and far of sound localization. *Current Biology* [online]. 1999, vol. 9, no. 10, pp. R361–R363 [accessed. 19. December 2013]. ISSN 09609822. Available from: doi:10.1016/S0960-9822(99)80227-9

MUSICANT, A.D. and R.A. BUTLER, 1985. Influence of monaural spectral cues on binaural localization. *Journal of the Acoustical Society of America*. 1985, vol. 77, pp.

202–208.

ORUÇ, İpek, Laurence T. MALONEY and Michael S. LANDY, 2003. Weighted linear cue combination with possibly correlated error. *Vision Research* [online]. 2003, vol. 43, no. 23, pp. 2451–2468 [accessed. 13. December 2013]. ISSN 00426989. Available from: doi:10.1016/S0042-6989(03)00435-8

RADEAU, M. and P. BERTELSON, 1977. Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Perception and Psychophysics*. 1977, vol. 22, pp. 137–146.

RAKERD, Brad, 1985. Localization of sound in rooms, II: The effects of a single reflecting surface. *The Journal of the Acoustical Society of America* [online]. 1985, vol. 78, no. 2, p. 524 [accessed. 24. July 2014]. ISSN 00014966. Available from: doi:10.1121/1.392474

RAKERD, Brad, 1986. Localization of sound in rooms, III: Onset and duration effects. *The Journal of the Acoustical Society of America* [online]. 1986, vol. 80, no. 6, p. 1695 [accessed. 24. July 2014]. ISSN 00014966. Available from: doi:10.1121/1.394282

RAYLEIGH, L, 1875. On our perception of the direction of a source of sound. *Proceedings of the Musical Association* [online]. 1875, pp. 75–84 [accessed. 16. April 2012]. Available from: http://www.jstor.org/stable/10.2307/765209

RECANZONE, Gregg H, 1998. Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proceedings of the National Academy of Sciences of the United States of America* [online]. 1998, vol. 95, no. 3, pp. 869–75 [accessed. 26. November 2014]. ISSN 0027-8424. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=33810&tool=pmcentrez&rendertype=abstract

RECANZONE, Gregg H, 2009. Interactions of auditory and visual stimuli in space and time. *Hearing research* [online]. 2009, vol. 258, no. 1-2, pp. 89–99 [accessed. 16. March 2012]. ISSN 1878-5891. Available from: doi:10.1016/j.heares.2009.04.009

ROSAS, Pedro, Johan WAGEMANS, Marc O ERNST and Felix A WICHMANN, 2005. Texture and haptic cues in slant discrimination: reliability-based cue weighting without statistically optimal cue combination. *Journal of the Optical Society of America. A, Optics, image science, and vision* [online]. 2005, vol. 22, no. 5, pp. 801–809. ISSN 1084-7529. Available from: doi:10.1364/JOSAA.22.000801

ROSAS, Pedro, Felix A WICHMANN and Johan WAGEMANS, 2007. Texture and object motion in slant discrimination: failure of reliability-based weighting of cues may be evidence for strong fusion. *Journal of vision* [online]. 2007, vol. 7, no. 6, p. 3. ISSN 1534-7362. Available from: doi:10.1167/7.6.3

ROSAS, Pedro and Felix A. WICHMANN, 2011. Cue Combination: Beyond Optimality. In: Julia TROMMERSHÄUSER, Konrad KORDING and Michael S. LANDY, eds.

Sensory Cue Integration [online]. B.m.: Oxford University Press, p. 144–152. ISBN 9780195387247. Available from: doi:10.1093/acprof:oso/9780195387247.003.0008

SANDERS, Lisa D, Benjamin H ZOBEL, Richard L FREYMAN and Rachel KEEN, 2011. Manipulations of listeners' echo perception are reflected in event-related potentials. *The Journal of the Acoustical Society of America* [online]. 2011, vol. 129, no. 1, pp. 301–9 [accessed. 18. December 2013]. ISSN 1520-8524. Available from: doi:10.1121/1.3514518

SEEBER, B., 2002. A new method for localization studies. *Acustica*. 2002, vol. in press.

SEIFRITZ, E., J.G. NEUHOFF, D. BILECEN, K. SCHEFFLER, H. MUSTOVIC, H. SCHACHINGER, R. ELEFANTE and F. DI SALLE, 2002. Neural processing of auditory looming in the human brain. *Curr Biol* [online]. 2002, vol. 12, no. 24, pp. 2147–51. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12498691

SEITZ, Aaron and Takeo WATANABE, 2005. A unified model for perceptual learning. *Trends in cognitive sciences* [online]. 2005, vol. 9, no. 7, pp. 329–34 [accessed. 20. July 2011]. ISSN 1364-6613. Available from: doi:10.1016/j.tics.2005.05.010

SEYDELL, Anna, David C. KNILL and Julia TROMMERSHÄUSER, 2011. Priors and Learning in Cue Integration. In: *Sensory Cue Integration* [online]. B.m.: Oxford University Press, p. 155–172. ISBN 9780199600458. Available from: doi:10.1093/acprof:oso/9780195387247.003.0009

SHAMS, Ladan and Aaron R SEITZ, 2008. Benefits of multisensory learning. *Trends in cognitive sciences* [online]. 2008, vol. 12, no. 11, pp. 411–7 [accessed. 21. January 2014]. ISSN 1364-6613. Available from: doi:10.1016/j.tics.2008.07.006

SHINN-CUNNINGHAM, B.G., 2000a. Adapting to remapped auditory localization cues: a decision-theory model. *Perception & psychophysics* [online]. 2000, vol. 62, no. 1, pp. 33–47. ISSN 0031-5117. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10703254

SHINN-CUNNINGHAM, B.G., N.I. DURLACH and R.M. HELD, 1998. Adapting to supernormal auditory localization cues I: Bias and resolution. *Journal of the Acoustical Society of America*. 1998, vol. 103, no. 6, pp. 3656–3666.

SHINN-CUNNINGHAM, B.G., S. SANTARELLI and N. KOPCO, 2000. Distance perception of nearby sources in reverberant and anechoic listening conditions: Binaural vs. monaural cues. In: . p. 88.

SHINN-CUNNINGHAM, Barbara B.G., 2000b. Learning Reverberation: Considerations for Spatial Auditory Displays. In: *Proceedings of the 2000 International Conference on Auditory Display* [online]. p. 126–134. Available from: http://www.icad.org/websiteV2.0/Conferences/ICAD2000/PDFs/ShinnCFinalNoFonts.p

df

SHINN-CUNNINGHAM, Barbara G., Norbert KOPCO and Tara J. MARTIN, 2005. Localizing nearby sound sources in a classroom: Binaural room impulse responses. *The Journal of the Acoustical Society of America* [online]. 2005, vol. 117, no. 5, p. 3100 [accessed. 18. July 2011]. ISSN 00014966. Available from: doi:10.1121/1.1872572

SCHOOLMASTER, M., N. KOPCO and B.G. SHINN-CUNNINGHAM, 2003. Effects of reverberation and experience on distance perception in simulated environments. *Journal of the Acoustical Society of America* [online]. 2003, vol. 113, p. 2285. ISSN 00014966. Available from: doi:10.1121/1.4780593

SCHOOLMASTER, Matthew, Norbert KOPČO and Barbara G. SHINN-CUNNINGHAM, 2004. Auditory distance perception in fixed and varying simulated acoustic environments. *The Journal of the Acoustical Society of America* [online]. 2004, vol. 115, no. 5, p. 2459 [accessed. 29. January 2014]. ISSN 00014966. Available from: doi:10.1121/1.4782332

SLUTSKY, D.A. and G.H. RECANZONE, 2001. Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*. 2001, vol. 12, no. 1, pp. 7–10.

SPEIGLE, J.M. and J.M. LOOMIS, 1993. Auditory distance perception by translating observers. In: . San Jose, California: IEEE Computer Society, p. 92–99.

STRYBEL, T.Z. and D.R. PERROTT, 1984. Discrimination of relative distance in the auditory modality: The success and failure ofthe loudness discrimination hypothesis. *Journal of Acoustic Society of America*. 1984, no. 76, pp. 318–320.

TAO, Qian, Chetwyn C H CHAN, Yue-Jia LUO, Jian-Jun LI, Kin-Hung TING, Jun WANG and Tatia M C LEE, 2013. How Does Experience Modulate Auditory Spatial Processing in Individuals with Blindness? *Brain topography* [online]. 2013, pp. 506–519. ISSN 1573-6792. Available from: doi:10.1007/s10548-013-0339-1

THOMPSON, Kelsey R., Daniel J. SANCHEZ, Abigail H. WESLEY and Paul J. REBER, 2014. Ego Depletion Impairs Implicit Learning. *PLoS ONE* [online]. 2014, vol. 9, no. 10, p. e109370. ISSN 1932-6203. Available from: doi:10.1371/journal.pone.0109370

TROMMERSHÄUSER, Julia, Konrad KORDING and Michael S. LANDY, 2011. *Sensory Cue Integration* [online]. B.m.: Oxford University Press. ISBN 9780195387247. Available from: doi:10.1093/acprof:oso/9780195387247.001.0001

UENO, Kanako, Norbert KOP, Barbara SHINN-CUNNINGHAM, Norbert KOPČO and Barbara SHINN-CUNNINGHAM, 2005. Calibration of speech perception to room reverberation. In: *Forum Acusticum*.

WARREN, R M, 1999. *Auditory Perception - A new Analysis and Synthesis*. Cambridge, UK: Cambridge University Press.

WEINBERGER, Norman M., 2015. *New perspectives on the auditory cortex* [online]. 1st ed. B.m.: Elsevier B.V. ISBN 1949824551. Available from: doi:10.1016/B978-0-444-62630-1.00007-X

WISNIEWSKI, Matthew G., Eduardo MERCADO, Klaus GRAMANN and Scott MAKEIG, 2012. Familiarity with speech affects cortical processing of auditory distance cues and increases acuity. *PLoS ONE* [online]. 2012, vol. 7, no. 7. ISSN 19326203. Available from: doi:10.1371/journal.pone.0041025

WISNIEWSKI, Matthew G., Eduardo MERCADO, Barbara a. CHURCH, Klaus GRAMANN and Scott MAKEIG, 2014. Brain dynamics that correlate with effects of learning on auditory distance perception. *Frontiers in Neuroscience* [online]. 2014, vol. 8, no. December, pp. 1–15. ISSN 1662-453X. Available from: doi:10.3389/fnins.2014.00396

WOODS, T.M. and G.H. RECANZONE, 2004. Visually Induced Plasticity of Auditory Spatial Perception in Macaques. *Current Biology*. 2004, vol. 14, pp. 1559–1564.

WOZNY, David R and Ladan SHAMS, 2011a. Computational characterization of visually induced auditory spatial adaptation. *Frontiers in integrative neuroscience* [online]. 2011, vol. 5, no. November, p. 75 [accessed. 16. June 2014]. ISSN 1662-5145. Available from: doi:10.3389/fnint.2011.00075

WOZNY, David R and Ladan SHAMS, 2011b. Recalibration of auditory space following milliseconds of cross-modal discrepancy. *The Journal of neuroscience : the official journal of the Society for Neuroscience* [online]. 2011, vol. 31, no. 12, pp. 4607–12 [accessed. 19. August 2011]. ISSN 1529-2401. Available from: doi:10.1523/JNEUROSCI.6079-10.2011

WRIGHT, B A and M B FITZGERALD, 2001. Different patterns of human discrimination learning for two interaural cues to sound-source location. *Proceedings of the National Academy of Sciences* [online]. 2001, vol. 98, no. 21, pp. 12307–12312. ISSN 0027-8424. Available from: doi:10.1073/pnas.211220498

WRIGHT, Beverly A and Yuxuan ZHANG, 2006. A review of learning with normal and altered sound-localization cues in human adults. *International journal of audiology* [online]. 2006, vol. 45 Suppl 1, no. Supplement 1, pp. S92–8 [accessed. 13. January 2014]. ISSN 1499-2027. Available from: doi:10.1080/14992020600783004

ZAHORIK, P., 1996. *Auditory Distance Perception: A Literature Review*.

ZAHORIK, P., 2002a. Direct-to-reverberant energy ratio sensitivity. *The Journal of the Acoustical Society of America* [online]. 2002, vol. 112, no. 5 Pt 1, pp. 2110–7 [accessed. 21. August 2011]. ISSN 0001-4966. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12430822

ZAHORIK, P., 2003. Auditory and visual distance perception: The proximity image effect revisited. *J Acoust Soc Am* [online]. 2003, vol. 113, no. 4, pp. 2270–2270. Available

from: link.aip.org/link/jasman/v113/i4/p2270/s5

ZAHORIK, Pavel, 2001. Estimating sound source distance with and without vision. *Optometry and vision science : official publication of the American Academy of Optometry* [online]. 2001, vol. 78, no. 5, pp. 270–5 [accessed. 27. January 2014]. ISSN 1040-5488. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11384003

ZAHORIK, Pavel, 2002b. Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America* [online]. 2002, vol. 111, no. 4, p. 1832 [accessed. 13. January 2014]. ISSN 00014966. Available from: doi:10.1121/1.1458027

ZAHORIK, Pavel, 2009. Perceptually relevant parameters for virtual listening simulation of small room acoustics. *The Journal of the Acoustical Society of America* [online]. 2009, vol. 126, no. 2, pp. 776–91 [accessed. 13. December 2013]. ISSN 1520-8524. Available from: doi:10.1121/1.3167842

ZAHORIK, Pavel and Paul W ANDERSON, 2014. 3aPPb13 . The role of amplitude modulation in auditory distance perception. 2014, p. 40292.

ZAHORIK, Pavel, Philbert BANGAYAN, V. SUNDARESWARAN, Kenneth WANG and Clement TAM, 2006. Perceptual recalibration in human sound localization: Learning to remediate front-back reversals. *The Journal of the Acoustical Society of America* [online]. 2006, vol. 120, no. 1, p. 343 [accessed. 16. April 2012]. ISSN 00014966. Available from: doi:10.1121/1.2208429

ZAHORIK, Pavel, Douglas S. BRUNGART and Adelbert W. BRONKHORST, 2005. Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica*. 2005, vol. 91, no. 3, pp. 409–420.

ZAHORIK, Pavel and Frederic L WIGHTMAN, 2001. Loudness constancy with varying sound source distance. *Nature neuroscience* [online]. 2001, vol. 4, no. 1, pp. 78–83. ISSN 1097-6256. Available from: doi:10.1038/82931

ZWIERS, M.P., A.J. VAN OPSTAL and G.D. PAIGE, 2003. Plasticity in human sound localization induced by compressed spatial vision. *Nat Neurosci* [online]. 2003, vol. 6, no. 2, pp. 175–81. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12524547

Appendix A

Despite the instructions, some of the subjects could not ignore sound level when judging distance as seen on **Figure A-1.** It shows the amount of correlation between perceived distances and level rove applied in R runs in the testing sessions 1, 5, 9, and 10. From the visual inspection of **Figure A-1** it could be noted that most of the subjects manage to follow the instructions but some subjects had high correlation with the intensity in the first session and improved in the following session, and few of the subjects did not improve at all. Consequently, we had a suspicion that the change in the strategy of responding was responsible for the reported learning effects.

The factorial analysis repeated measures ANOVA with similar design as the one in the main analysis (factors: session, run, condition, init group) was run only for subjects who ignored level. The ANOVA included 22 subjects, the factor init group was imbalanced, the procedure of software CLEAVE (Herron 2005) allows to compute the F statistics also for the partially imbalanced design. The result showed the main effect of condition ($F(1,20)=25.38$, $p<0.01$), the interaction of init group x session ($F(2,40)=12.56$, $p<0.05$), and the interaction of session x condition ($F(2,40)=4.62$, $p<0.05$). The main trends in data were preserved except the interaction with factor run.

Figure A-2 shows the learning effects of the two training regimens in the two testing conditions for the subjects who ignored level of presentation. Separate repeated measures ANOVA showed the main effect of testing condition ($F(1,21)=4.82$, $p<0.05$) similarly to the main analysis but the interaction was not preserved. An additional t-test was performed on the averaged learning data (average of four columns of **Figure A-2**) to assess whether the total learning differed from zero. The result showed non-significant difference from zero (t-test: $p>0.1$), which suggests that the loudness had a potential to interfere to a certain extent with the current findings. However, the main trends were preserved and the result could relate to the fact that 10 subjects who started with lower performance were excluded.

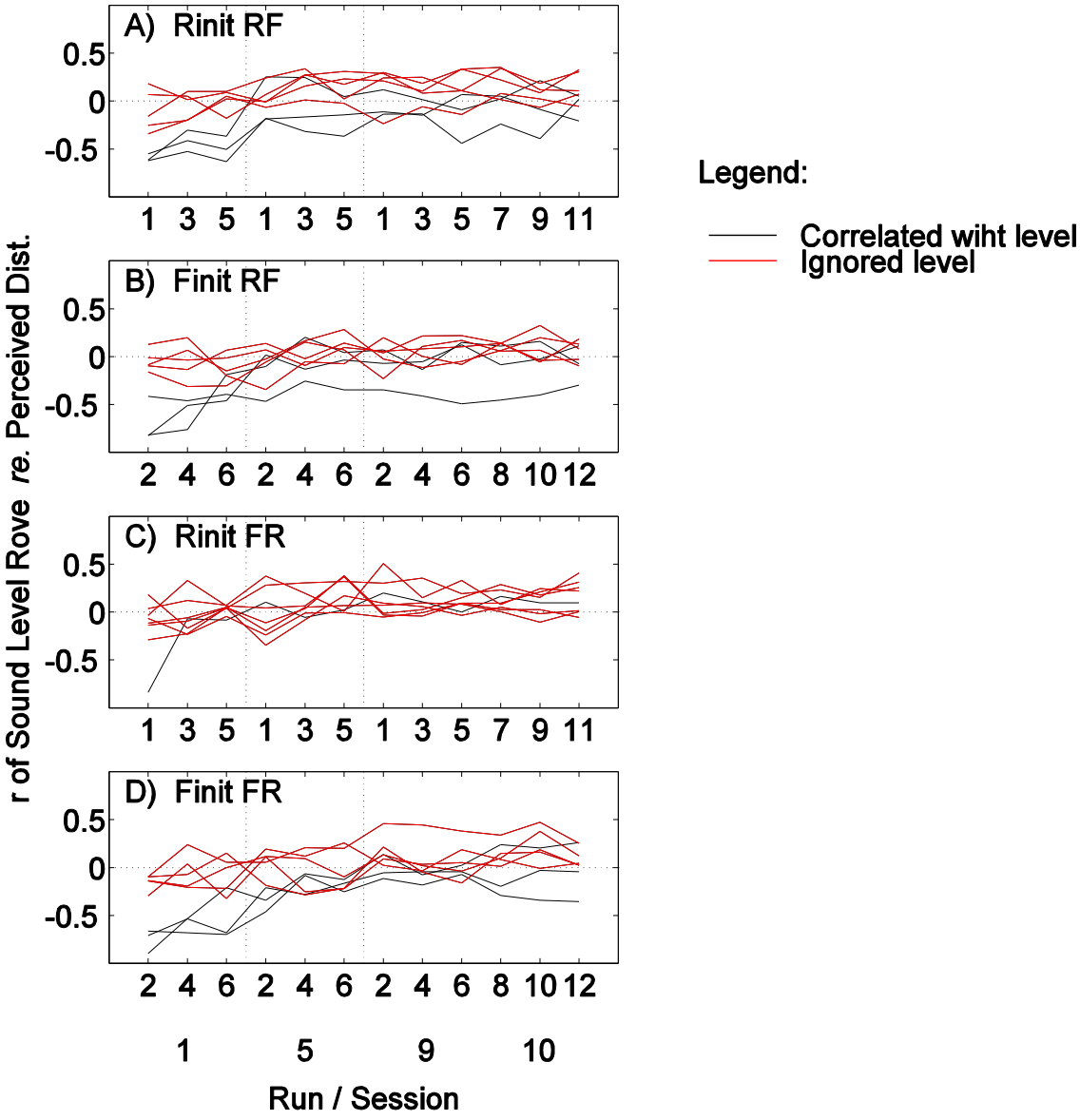

Figure A-1 Correlation of responses and level rove during testing runs in which the level of presentation was roved on trial-by-trial basis (R runs). X-axis shows run number in corresponding testing session. Each line shows data of one subjects, divided by experimental groups form (A)-(D). Black lines are data of subjects that had the highest correlation at the beginning of the experiment. Subjects who exceeded correlation 0.4 during the first R run (black lines) were excluded, while the rest was used in subsequent ANOVA to control for the effect of level presentation on learning.

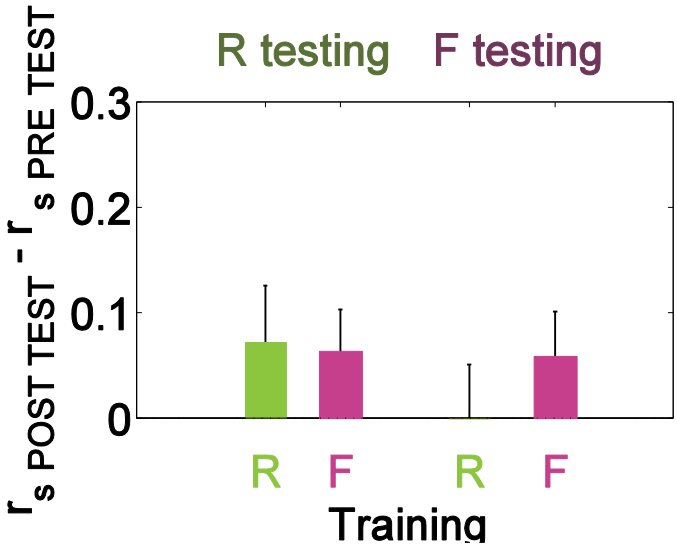

Figure A-2 the same caption as Figure 2-9 but the data show only subjects who could ignore sound level (red lines on Figure A-1).

Appendix B

The following two figures list the localization performance in terms of response bias (**Figure B-1)** and repose standard deviation (**Figure B-2**) of audio-visual Experiment 2 as the raw data. The general trends of performance of Experiment 2 were similar to Experiment 1, which was shown in the main text, therefore these figures were omitted from the results section.

Figure B-3 shows the variability accounted by the power model of the auditory distance perception (Zahorik et al. 2005) measured by correlation coefficient. In comparison with a similar study (Anderson and Zahorik 2014), the responses in the current experiment were less precise; however, current findings support the previously observed difference between the AV and A trials (it also supports the analysis of SDs). In addition to that, the figure shows the difference of correlation coefficients between the A V-Closer and A V-Farther presentations in Experiment 2, which was observed in the analysis of SDs. Furthermore, the figure also shows the comparison of the A responses in the Experiment 3 with the A responses in the main experiments 1 and 2. The statistical analysis showed a significant difference between the A trials in the control experiment and the A trials in the Experiment 1 and Experiment 2 even if the analysis of SDs did not show this difference. The performance in the control experiment was worse than in the main experiments. The change in response precision can however relate to change in compression which may have been more pronounced in the experiments which involved the AV training.

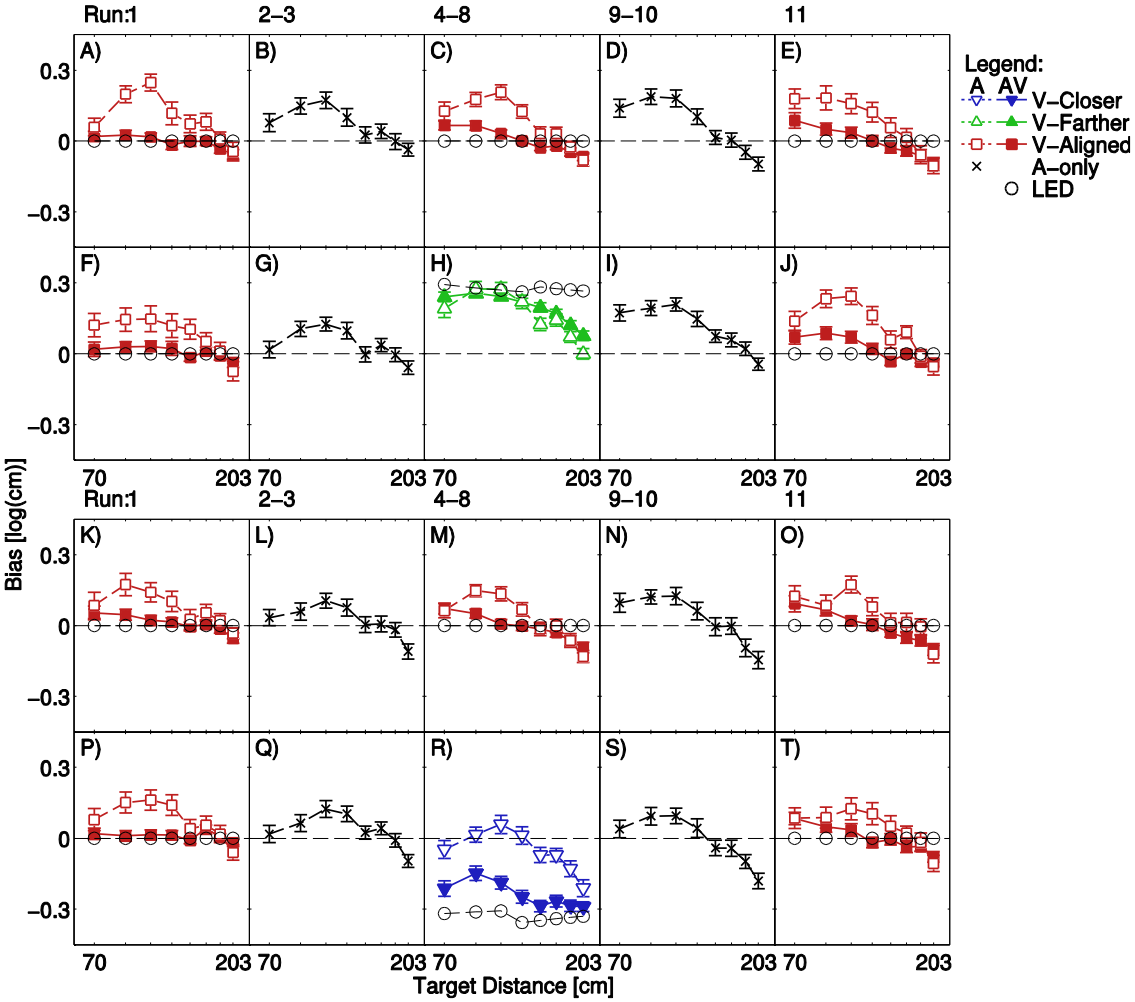

Figure B-1 Experiment 2 localization bias. (A-J) Data of subject group V-Farther, V-Aligned, (K-T) data of subject group V-Closer, V-Aligned. The figure layouts of the two upper rows (A-J) and bottom two rows (K-T) are identical to layout of Figure 3-3. The rows stand for sessions, the columns divide the experimental session with the pattern that was used in the main analysis. Open symbols represent A stimuli, closed symbols AV stimuli. See legend for the color coding.

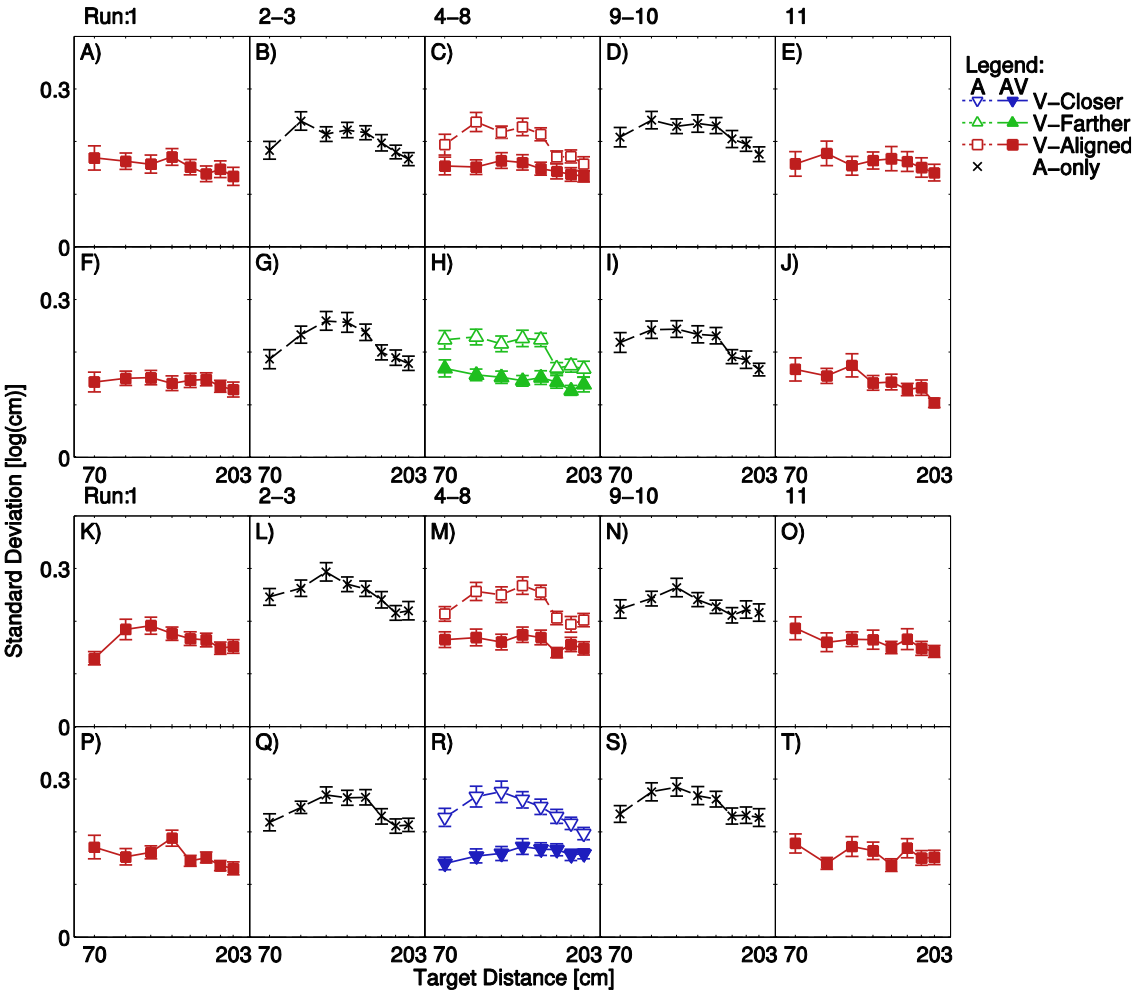

Figure B-2 Experiment 2 response SDs. The data were computed exactly in the same was as data in Experiment 1. The figure is organized with the same layout as the Figure B-1.

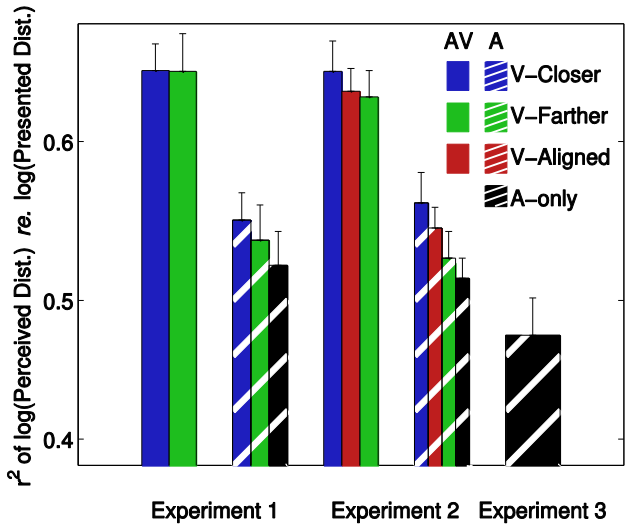

Figure B-3 Square of Pearson's correlation coefficients of the perceived *re.* presented distance averaged across adaptation runs (4:8) in three experiments. Data

also express the variance accounted by the power model (Anderson and Zahorik 2014). The results of the statistical analysis RM ANOVA are similar to the results of the SDs in terms of main effects and interactions (Sec. 3.5.2). However, statistical comparison of the Experiment 3 data and the data of the other two experiments shows a significant difference (Welch's t-test: Exp. 2 + Exp. 3 vs. Exp. 4, $p<0.05$).