Auditory Cues in the Perception of Self Motion

B. Kapralos[1,4], D. Zikovitz[2,4], M. Jenkin[1,4] and L. R. Harris[2,3,4]

[1]Dept. of Computer Science, York University. North York Ontario, Canada. M3J 1P3
[2]Dept. of Biology, York University. North York Ontario, Canada. M3J 1P3
[3]Dept. of Psychology, York University. North York Ontario, Canada. M3J 1P3
[4]Centre for Vision Research, York University, North York Ontario, Canada. M3J 1P3

Acknowledgements:

- The help of Konstantinos Derpanis is gratefully acknowledged. He provided ideas and insight throughout the entire process and was the first to experiment with any of our test iterations.

Overview (1):

- **Introduction**
  - Motivation
    - Importance of spatial sound
    - Self-motion perception
  - Previous work related to self-motion perception
  - Project goals
- **Background**
  - Auditory distance perception
  - Auditory motion perception

Overview (2):

- **Methods**
  - Experimental overview
  - Apparatus
  - Experimental details
- **Results**
  - Summary of experimental results
- **Conclusions**
  - Overall summary
  - Future work

Introduction

Motivation (1):

- **Hearing Provides Info About our Environment**
  - Spatial sounds give detailed info of our surroundings
    - Determine direction and distance to objects
    - Warn of approaching dangers e.g. predators
  - Unlike vision, hearing is omni-directional
    - Can hear sounds in back of us
    - Can hear in complete darkness
  - Can guide the more "finely tuned" visual attention
    - Eases the burden of the visual system

Motivation (2):

- **Importance of Spatial Audio in a VE**
  - Maintains sense of environmental realism
    - Helps increase immersion and hence presence
  - Can enhance perception of video quality
- **Spatial Audio Often Ignored in a VE**
  - Emphasis typically placed on visual cues (graphics)
  - Poor cues which don't reflect natural spatial cues
  - Little emphasis on distance and motion simulation despite their potential to increase realism
    - Flight simulators etc.

Motivation (3):

- **Self-Motion Perception**
  - In our natural environment, using a variety of modalities, we can judge (estimate):
    - Our own self-motion
    - The motion of approaching or receding objects
  - This ability is beneficial and at times critical to our survival
  - Developing an understanding of mechanisms and modalities responsible for this can lead to more accurate simulations

Motivation (4):

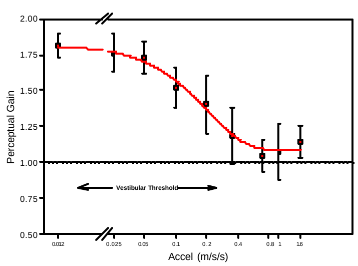
- **Self-Motion Perception (cont.)**
  - Majority of studies have focused primarily on vision and vestibular cues
    - We use visual and vestibular information to gauge self-motion
  - Despite importance of auditory cues in the understanding of our environment
    - Very little is known with respect to their contribution and interaction with other cues, in self motion estimation

Previous Work (1):

- **Harris, Jenkin & Zikovitz (2000)**
  - The "roles of visual & vestibular cues in determining perceived distance of passive linear self motion"
  - Vestibular only:
    a. Subject shown target
    b. Subject blindfolded & accelerated towards target
    c. Subject indicated when target reached
  - Subjects over-estimated target distance
    - Subjects felt they traveled further than they actually had

Previous Work (2):

- **Redlick, Jenkin and Harris (2001) - Vision Only**
  - Over-estimation of motion
  - Increasing acceleration ? increasing accuracy



Project Goals:

- **Investigate Auditory Self-Motion Perception**
  - Build on work of Harris et al.
    - Develop a clearer understanding of auditory motion cues and their interaction with vestibular cues
    - Examine effect auditory cues have when available alone or in conjunction with vestibular cues
    - What are the consequences of these effects with respect to a "multi-modality" virtual environment ?

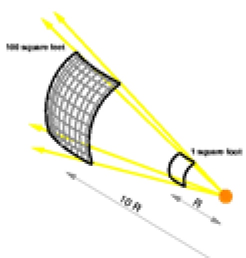
Background: Auditory Distance & Motion Perception

Auditory Distance Perception (1):

- **Relative & Absolute Distance Cues**
- **Main Auditory Distance Cues**
  - Sound level (intensity)
  - Reverberation
  - Sound source frequency spectrum
  - Binaural (interaural time and level differences - ITD and ILD)
  - Sound source spectrum characteristics (content)

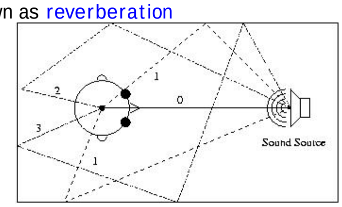
Auditory Distance Perception (2):

- **Sound Level (Intensity)**
  - Decrease of sound intensity as distance increases
    - e.g. inverse square law (approx. - but not quite!)



Auditory Distance Perception (3):

- **Reverberation**
  - Collection of reflected waves (possibly 1000s) is known as reverberation



  - Ratio of direct-to-reverberant energy is a powerful cue to sound source distance

Bill Kapralos
Center for Vision Research October 17 2003

Auditory Distance Perception (4):

- **ILD and ITD as a function of distance**
  - ILD highly dependent on distance for sound source in the near-field (within ~0.5m)
    - Especially, when source is away from median plane
  - ITD dependence is much smaller
- **Source Spectral Content**
  - Attenuation of high frequency components
  - Primarily for large distances (> 15m)
  - "Weak" cue

Auditory Distance Perception (5):

- **Auditory Distance Estimation Summary**
  - Intensity and ratio of direct-to-reverberant energy are primary cues
  - Intensity alone is a poor distance cue!
    - Relative cue only!
    - Leads to under-estimation of source distance
  - Distance judgments are much more accurate under reverberant conditions
    - Can potentially provide absolute distance judgment

Auditory Motion Perception (1):

- **Four Main Auditory Motion Cues**
  1. Sound level (intensity) changes
  2. Interaural temporal and amplitude differences (binaural cues)
     - Primarily for horizontal motion
  3. Doppler frequency shift
     - Primarily for higher velocities
  4. Reverberation

Auditory Motion Perception (2):

- **Very Little Research Related to Dynamic Sound Localization**
  - Sound localization studies have focused primarily on static sound source and static observer
  - Majority of motion studies have examined velocity discrimination
    - Horizontally moving sound source
  - Acoustical time-to-contact
    - Echolocation

Auditory Motion Perception (3):

- **Very Little Agreement on How Auditory System Encodes a Moving Sound Source**
  - Do we posses “special” auditory motion detectors ?
- **Two “Theories” for Dynamic Sound Localization**
  - Snap-shot hypothesis
  - Specialized motion detectors
  - Perhaps we employ both mechanisms?

Auditory Motion Perception (4):

- **Snap-Shot Hypothesis (Grantham 1986)**
  - Mechanisms responsible for static sound localization are used for auditory motion detection
  - “Sample” auditory space at discrete times – each sampling gives one “snap-shot”
    - Integrate snap-shots over time interval to detect moving sound
  - Cannot determine velocity/acceleration directly
    - Calculated using total distance traveled and time

Auditory Motion Perception (5):

- **Specialized Motion Detectors**
  - Mechanisms in the auditory system that respond directly to the motion of auditory targets
  - May respond to auditory targets moving in specific directions or particular velocities
  - No “compelling physiological evidence of such a mechanism” (Perrot et. al. 1993) but
    - Several studies indicate such “units” in the cat and monkey auditory system (Ahissar et al, 1992, Yin and Kuwada 1983)

Auditory Motion Perception (6):

- **Auditory Motion Estimation Summary:**
  - Intensity is the primary cue
  - Auditory motion is over-estimated regardless of whether observer is passive or active
    - Accuracy improves when observer is active
    - Greater over-estimation when sound source is approaching as opposed to receding

Methods

Experimental Overview (1):

- **Four Experiments to Test Four Conditions**

1. Motion Only: Physical motion (constant acceleration) of subject
2. Motion + Audio: Physical motion (const. accel.) of subject with stationary sound source at starting position
3. Audio Only: Stationary subject and sound source but decreasing sound level (intensity)
4. Moving Audio: Stationary subject with sound source moving (const. accel.) away from subject

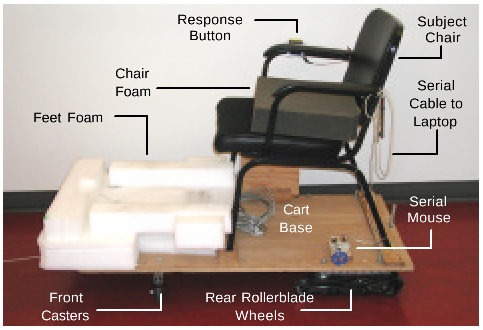
Experimental Overview (2):

- **Procedure**

1. Subject at starting position – physical target shown to them (target at 1m, 2m, 3m or 4m from subject)
2. Subject blindfolded
3. Presented with stimulus (auditory, vestibular or both)
4. Subjects indicated when they reached the target by pressing button
5. Subject brought back to starting position (when physical motion involved)

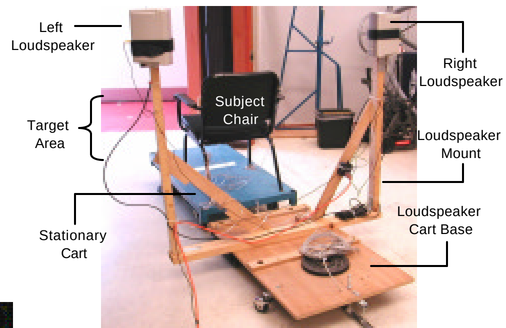
Apparatus - Subject Cart:

- **Subject Seated on Cart ? Conditions 1,2,3**



Apparatus – Loudspeaker Cart:

- **Loudspeaker Moved with Cart ? Condition 4**

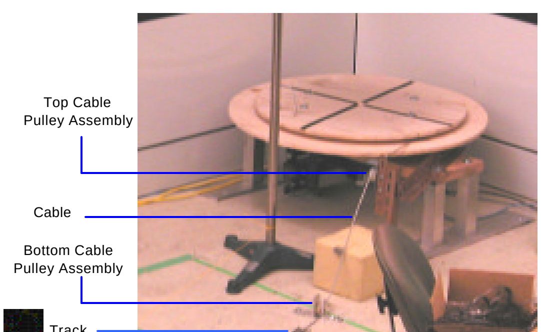


Apparatus - Motion Generation: (1):

- **Moving the Subject or Sound Source - Motor & Motor Controller**
  - Motor used to pull cart (via motor controller) according to one of five constant acceleration profiles:
    - 0.012ms^{-2} 0.025ms^{-2} 0.05ms^{-2} 0.1ms^{-2} or 0.2ms^{-2}
  - Cart forced to follow "track" placed on floor
    - Ensured straight line trajectory
  - Depending on condition being tested, subject or sound source placed on cart

Apparatus - Motion Generation (2):

- **Motor and Pulley Assemblies**



Apparatus - Auditory Stimulus (1):

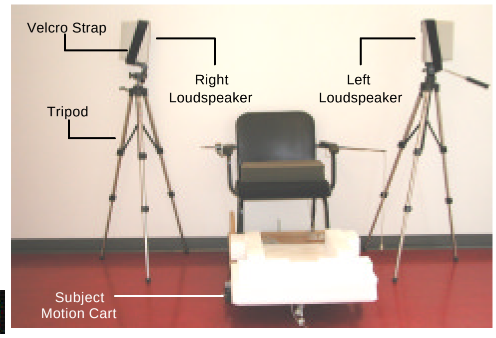
- **Characteristics**
  - Uniformly distributed "white" noise
  - Broadband: 200Hz – 10kHz
    - More accurate distance estimates as opposed to single tone stimulus
  - Three sound stimuli levels (66dB, 69dB, 72dB)
    - For each trial, level of stimulus randomly chosen
    - Minimize potential to associate sound level with particular distance or acceleration

Experimental Details (1):

- **Motion Only**
  - Repeat experiment performed by Harris et al.
    - Reference measure for other experiments
  - Vestibular cues only – no audio cues & motion in the dark
- **Audio + Motion**
  - Same as motion only condition except for stationary sound source at starting position
    - Sound source intensity constant but relative to subject, as they move away, intensity decreases
  - What effect will decreasing intensity have ?

Experimental Details (2):

- **Set-up for Conditions 1,2,3**



Experimental Details (3):

- **Audio Only**
  - No physical motion
    - Subject and sound source remained stationary
  - Sound source intensity decreased
    - Imitate reduction which would occur if subject was accelerated from stationary sound source according to one of the file acceleration profiles
  - Same auditory stimuli as in condition 1 – except for intensity decrease
  - Can reduction of source intensity provide a reliable cue to self-motion ?

Experimental Details (4):

- **Moving Audio Source**
  - Subject remained stationary at starting position while sound source accelerated away from them according to one of the five acceleration profiles
  - Since the source is actually moving, other auditory cues (e.g. reverberation) will be present
    - How will this affect results of “audio only” condition ?
  - Can a moving sound source induce the illusion of self-motion to a stationary subject ?
    - “Auditory vection”

Experimental Details (5):

- **Video Example**

Results

Perceptual Gain:

- **Perceptual Gain (g) to Measure Performance**

  Ratio between: $d_{perceived}$ to d_{actual}

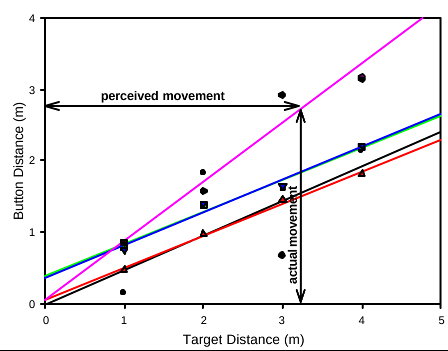
  d_{actual} ? distance subject pressed button

  $d_{perceived}$? target distance – distance subjects perceived they traveled

- Ideal situation ? unity gain
  - Gain > 1 ? over-estimation - subjects felt they traveled further than they actually did
  - Gain < 1 ? under-estimation – subjects felt they traveled less than they actually

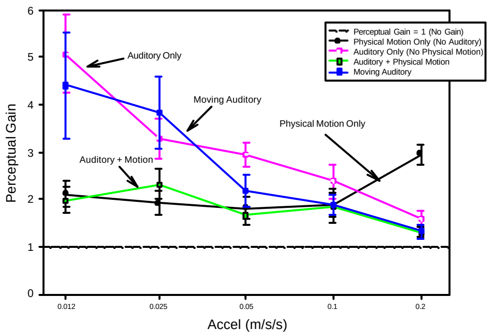
Perceptual Gain Derivation:

Audio + Motion ? One subject, five accelerations



Averaged Perceptual Gain:

Perceptual Gain by Acceleration



Summary of Results:

- **Experimental Conditions Significantly Different**
  - Repeated measures ANOVA
- **"Slow" Accelerations Significantly Different From "Fast" Accelerations**
  - Slow Accelerations ? 0.012, 0.025, 0.05
  - Fast Accelerations ? 0.1, 0.2

Results – Motion Only (1):

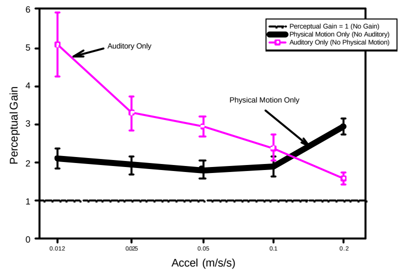
- **Motion Only**
  - Results of Harris et al. verified
    - With vestibular cues alone, subjects over-estimated self-motion systematically with acceleration

Results – Motion Only (2):

- **Vestibular Threshold**
  - Slower accelerations are below the reported vestibular threshold of ~ 0.014
  - Motion should not be detected directly by the vestibular system!
    - Not so! over-estimated but less than audio only
  - Other potential cues may lead to motion detection:
    - Initial motor jerk, wind, vibration, noise
    - Steps taken to reduce these effects but cannot eliminate completely

Results – Motion Only (3):

- **Motion Only**
  - Significantly different from audio only

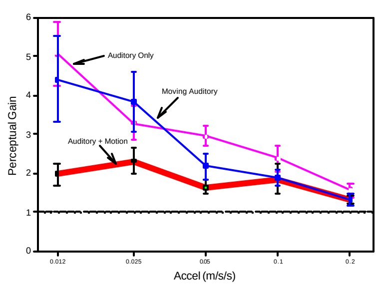


Results – Audio + Motion (1):

- **Audio + Motion**
  - Most accurate condition and significantly different from other conditions
  - Addition of auditory info. results in more accurate self-motion perception
  - Evidence indicating integration of multi -sensory info. ? more accurate judgments
  - Motion only and motion + audio fairly similar (to $0.1 ms^{-2}$)
    - Shift in performance between motion only and audio + motion after this point

Results – Audio + Motion (2):

- **Audio + Motion**
  - Significantly different from moving audio and audio only

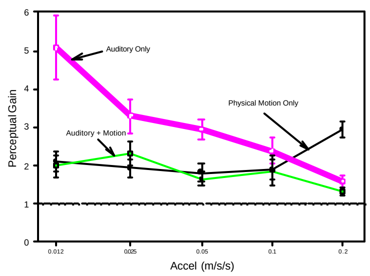


Results- Audio Only (1):

- **Audio Only**
  - Least accurate condition
    - More accurate for higher accelerations
    - To be expected – intensity alone is repeated to be a poor cue to source distance and motion (constant velocity)
    - Performance does increase with increasing acceleration in line with work of Perrot et al., 1992

Results – Audio Only(2):

- **Audio Only**
  - Significantly different from audio + motion, motion only

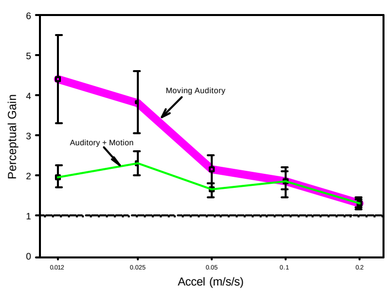


Results – Moving Audio (1):

- **Moving Audio**
  - No difference whether audio source moved away from subject or intensity of stationary source decreased
    - Over-estimation of self-motion
  - Would expect increased accuracy as more cues become available (e.g. reverb.) as source moves away
    - Is room too big for reverb cues to be useful ?
  - Significantly different from audio + motion
    - More accurate with two cues
  - No "auditory vection" induced by moving sound

Results – Moving Audio (2):

- **Moving Audio Only**
  - Significantly different from audio + motion



Conclusions

Conclusions (1):

- **Summary**

1. No difference between moving (receding) sound source and stationary sound source whose intensity is decreased
   - Is reverberation important ?
   - No auditory vection
2. Decreasing sound source intensity is an inaccurate cue to self-motion perception, particularly at low accelerations

Conclusions (2):

- **Summary (cont.)**

3. Self-motion estimation most accurate with auditory and vestibular cues as opposed to each cue alone
   - Below ~ $0.1 ms^{-2}$, physical motion captures the auditory stimulus but…
   - Auditory stimulus captures physical motion for higher accelerations
4. Physical motion is systematically over-estimated at both high and low accelerations

Conclusions (3):

- **Future Work**
  - Introduction of visual cues
    - Examine interaction between auditory and visual cues and auditory, visual and vestibular cues
    - Potentially useful for VR applications
  - Current experiments involved decreasing sound intensity - receding sound source
    - More accurate for judging motion/distance of approaching sound sources
    - How about increasing sound intensity - approaching sound source ?

Conclusions (4):

- **Future Work (cont.)**
  - "Auditory" vection – can a stationary subject be made to think they are moving?
    - Active array of loudspeakers moving past stationary subject
    - Stationary array of loudspeakers, stationary subject - sound applied to one loudspeaker at a time from starting position to end position
    - Possible ? revolving acoustical stimulation can produce "audio-kinetic" circular self-motion perception Von Stein, 1910, Dodge, 1923

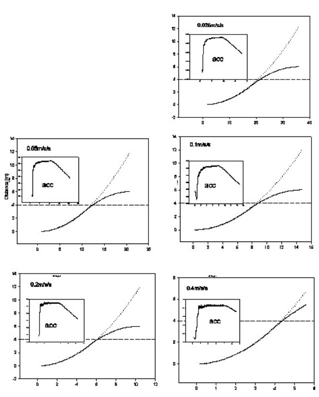
Finally… The End!

Questions ??

Motion Profiles:

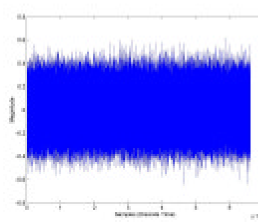
- **Motion Profiles**
  - Great effort to ensure accurate and repeatable profiles

Main plot:
x-axis ? time (s)
y-axis ? velocity
Inset:
x-axis ? time (s)
y-axis ? acceleration

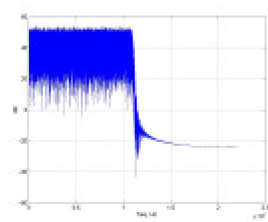


Auditory Stimulus:

- **Graphical representation**



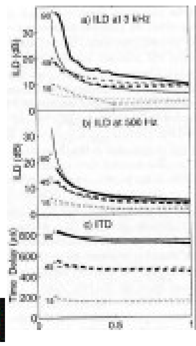
Time Domain



Fourier Domain

Auditory Distance Perception (i):

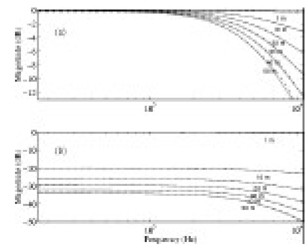
- **ILD and ITD as a function of distance**



- ILD highly dependent on distance for sound source in the near field (within 0.5m) especially, when source is away from median plane
- ITD dependence is much smaller

Auditory Distance Perception (ii):

- **Source Spectral Content**
  - Relative cue
  - High frequency attenuation for large distances (> 15m)
  - "Weak" cue



Averaged Perceptual Gain (1):

Condition	Perceptual Gain per Acceleration				
	$0.012ms^{-2}$	$0.025ms^{-2}$	$0.05ms^{-2}$	$0.1ms^{-2}$	$0.2ms^{-2}$
Motion Only	2.11	1.93	1.81	1.88	2.95
Audio + Motion	1.98	2.33	1.65	1.86	1.31
Audio Only	5.08	3.29	2.95	2.38	1.58
Moving Audio	4.41	3.84	2.18	1.89	1.32

Summary of Results (1):

- **Experimental Conditions (based on gain)**
  - Main effect among the four conditions
    - ANOVA: $\mathbf{F}(3, 40) = 10.22$, $p < 0.001$

Condition	Gain	Different From
Audio + Motion	1.83	Moving Audio, Audio Only
Motion Only	2.14	Audio Only
Moving Audio	2.68	Audio + Motion
Audio Only	3.06	Audio + Motion, Motion Only

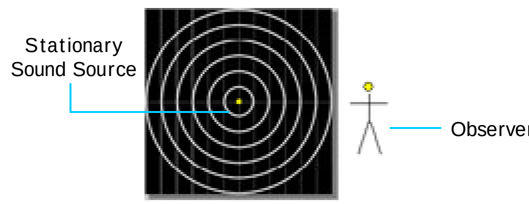
Summary of Results (2):

- **Effect of Acceleration (cont.)**
  - Summary of Tukey-Kramer comparison

Acceleration (ms^{-2})	Gain	Different From Acceleration
0.012	1.78	0.2, 0.1, 0.05
0.025	2.00	0.2, 0.1, 0.05
0.05	2.15	0.025, 0.012
0.1	3.00	0.025, 0.012
0.2	3.20	0.025, 0.12

Auditory Motion Perception (1):

- **Doppler Shifts – Stationary Sound Source**
  - Sound source propagates uniformly in all directions



Auditory Motion Perception (2):

- **Doppler Shifts – Moving Sound Source**
  - Change in wavelength since wave crests
    - "Bunch-up" in direction of movement
    - Spread out in opposite direction