Auditory Distance Perception

Yan-Chen Lu & Martin Cooke

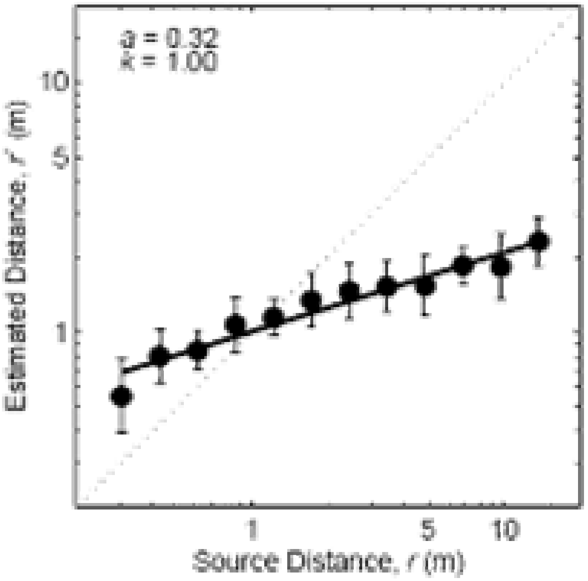
SPandH

The University Of Sheffield.

Human auditory distance perception

Human performance data (21 studies, 84 data sets) can be modelled by a power function $r' = kr^a$ (Zahorik et al. 2005)

r : physical source distance
r' : perceived distance
a : 0.15 ~ 0.70 (mu=0.54)
k : ~= 1.3



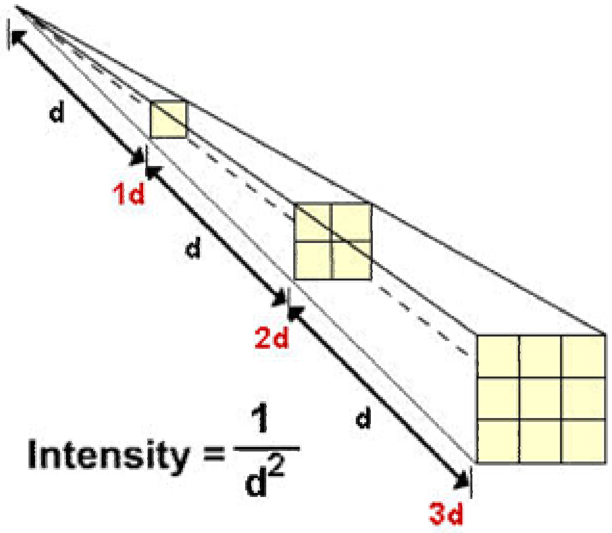
- Systematic bias: listeners significantly
  - **overestimate** distance to sources < 1m
  - **underestimate** the distance to sources > 1 m
- cf angular direction accuracy (1° frontal to 10° general)

Possible determinants of perceived distance

- Static acoustic cues
  - Intensity
  - Direct to reverberant energy ratio
  - Spectrum
  - Binaural cues
- Dynamic acoustic cues
  - Acoustic τ
  - Motion parallax
- Vision
- Familiarity

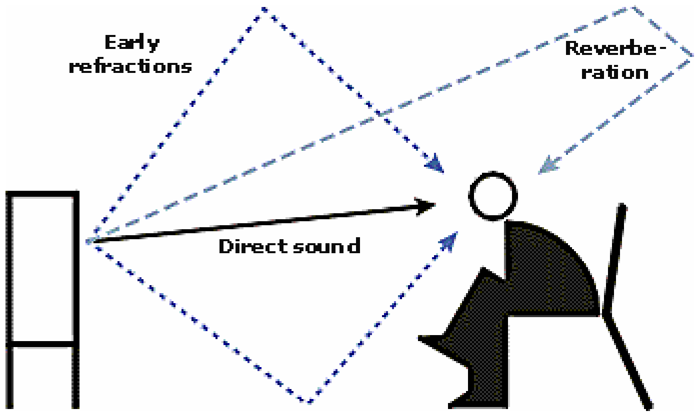
Intensity

- In free-field, intensity falls off by 6 dB for each doubling of distance
  - less in reverberant environments e.g. 4.25 dB in normal room (Zahorik & Wightman 2001)
- Could serve as an *absolute* distance cue, if source properties (acoustic power & radiation pattern) familiar to listeners
- More likely useful as a *relative* distance cue
- JNDs: ~10% change in distance, more for close sources



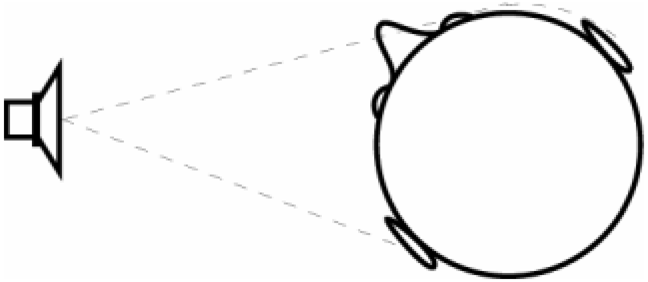
Direct to reverberant ratio

- Reverberant energy in enclosed space may be approx. uniform over varying source positions (diffuse environment)
- Hence, closer sources produce a greater ratio of direct-to-reverberant energy
- Serves as an *absolute* distance cue when listener is familiar with the environment
- Listeners can learn reverberation (Kopco et al., 2004)
- JND
  - 5~6 dB change in d-r energy ratio
  - equivalent to 100% distance change!
  - too large to be a *relative* distance cue



Spectral cues

- This cue is processed differently for near-field and far-field (> 15m) sources (Coleman, 1968)
  - Foghorn design principle: selective absorption of high frequencies (Blauert 1997)
  - Emphasis effect (near-field) for low-frequency energy from acoustic scattering by head and torso (Brungart et al. 1999)
  - role in ADP not yet clear



Dynamic cue: acoustic τ

- = motion-induced intensity rate of change
- Participants walked for 1.5m then asked to make a distance judgement (Ashmead et al., 1995)
- Better judgements than when standing still
- listeners underestimate τ by similar amount as distance (Guski, 1991)

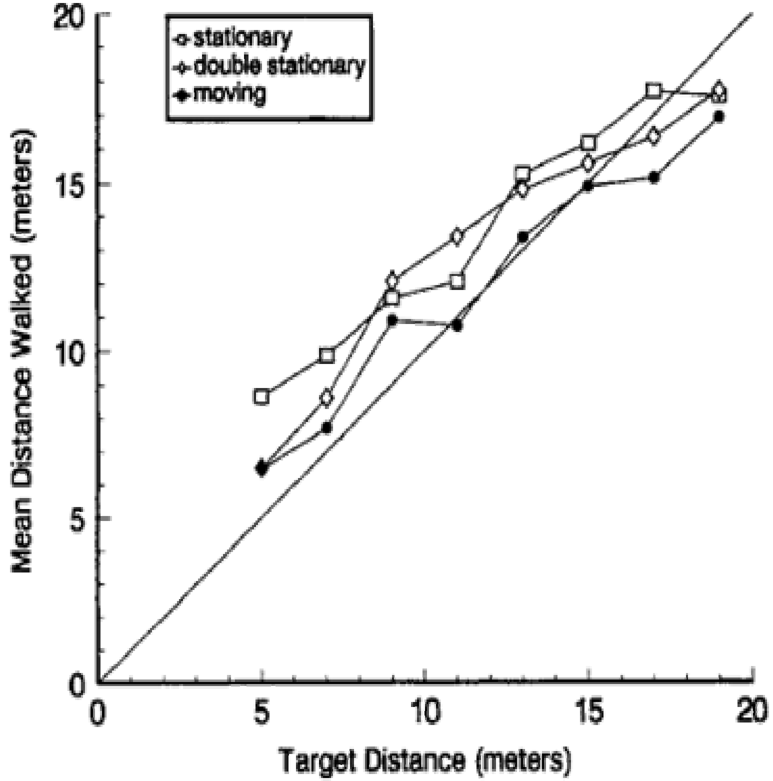
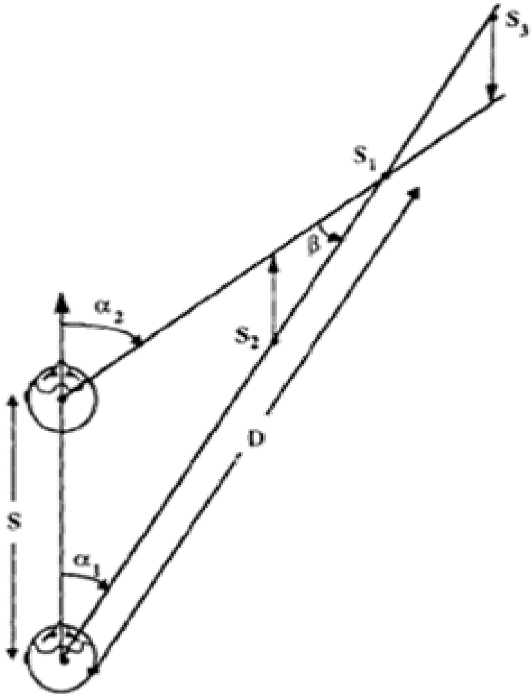


Figure 4. Mean distance walked as a function of target distance and listening condition (Experiment 2). The diagonal line corresponds to perfect performance.

Dynamic cue: motion parallax

- = change in angular direction of the sound source resulting from translation of source or listener
- *Marginally* improves distance perception accuracy (Speigle & Loomis 1993)
  - NB experiment used both acoustic τ and motion parallax



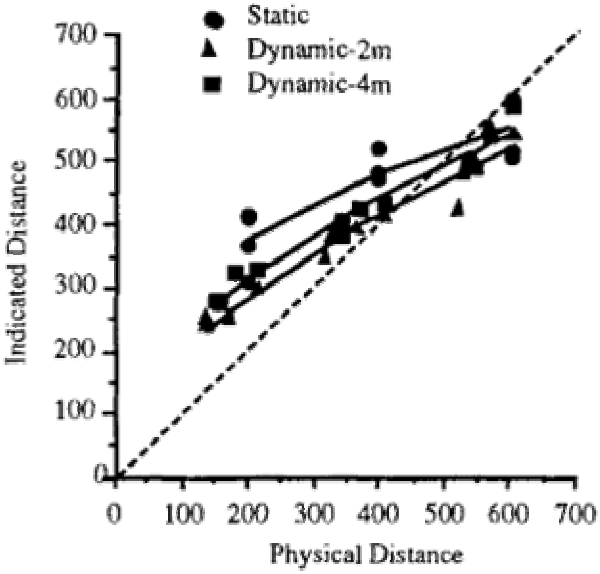


Figure 3. Mean indicated distance for the three conditions, averaged over source azimuth.

Familiarity

- Listeners rely on prior listening experience (Brungart & Scott 2001)
- Whispered speech always assumed to be nearby and shouted speech is far away

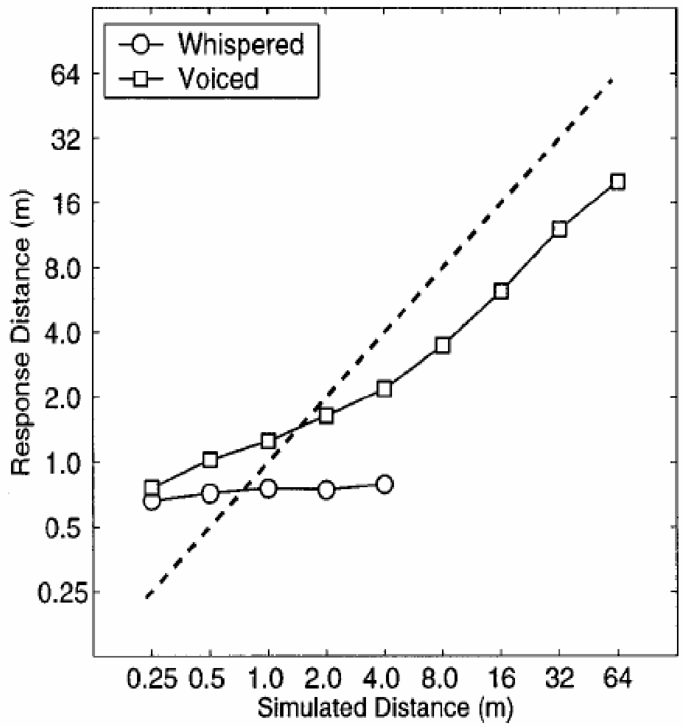


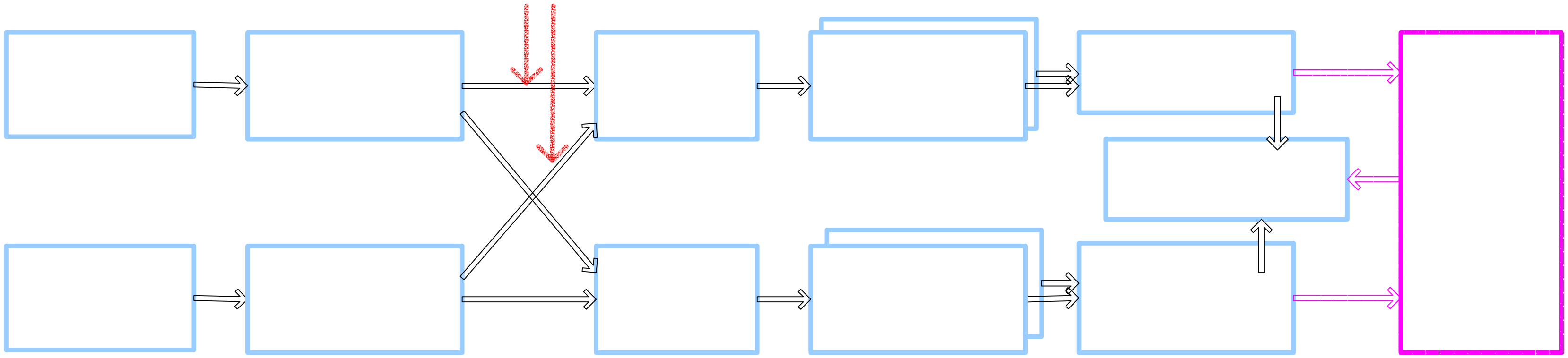
FIG. 4. Effects of voicing on the distance perception of speech in experiment 1. Data are similar to Fig. 3 except that the mean distance judgments have been plotted separately for the unvoiced trials where whispered speech was used as the stimulus and the voiced trials where conversational or shouted speech was used for the stimulus. Each data point represents the average response distance across all of the trials presented at that particular simulated distance with voiced speech or whispered speech. Note that the whispered speech was perceived at about 0.8 m across all the simulated distances.

ADP unknowns

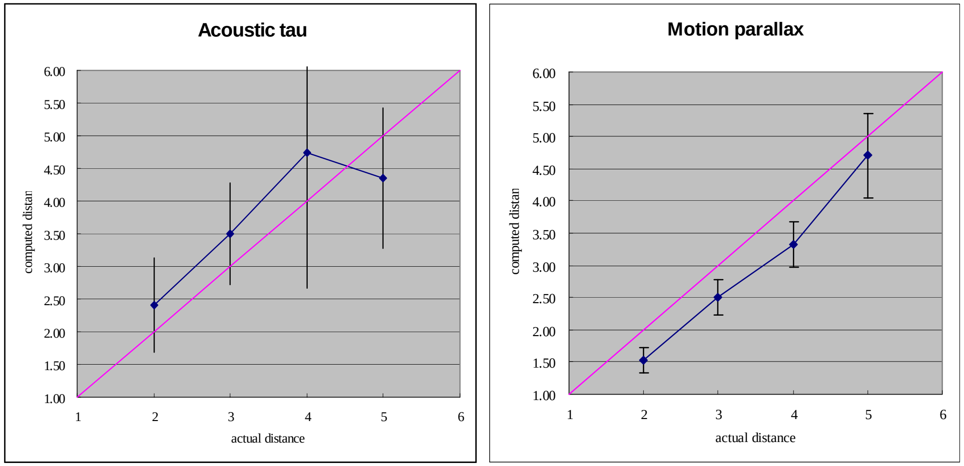
- cue weighting
- integration with visual cues
- integration with azimuth estimation
- relative roles of data and prior knowledge
- little known about noisy/multisource environments
- scope of listener-induced ie. active POP-like strategies
  - dynamic acoustic distance cue using translation or head rotation
  - turn to face target to improve azimuth estimation when incorporating motion parallax cue
  - turn 90 deg to tackle front-back confusions
- no general-purpose computational models

An initial model

- Use triangulation and inverse square law to estimate distance based on the ITD and intensity observation(s) over time



- simple 4-position motion path simulated in roomsim
- very early results (small data set)



Basilar M

Current & future work

- develop Bayesian framework for both static and active (azimuth,distance) estimation, incorporating multiple cues, prior experience, visual priors and online model estimation
- evaluation in source location and tracking scenario (POP demonstrator)
- possible collaborative behavioural study on ADP in multisource environments
- room for more neuroscience studies on multimodal distance maps?
  - role of right superior temporal cortex (for intensity-based distance cue; Mathiak et al., 2003)

Others

Vision

- “ventroloquism effect”: a visible target can capture the perceived location of sound source
  - up to 30deg in azimuth (Jack & Thurlow, 1973)
  - ?? for distance (Zahorik, 2001)

Binaural

- Interaural time difference (ITD) & level difference (ILD) are dependent on source distance within 1m (Brungart et al. 1999)
  - ILD increases rapidly for approaching source while ITD increases only modestly