Poster: Spatial Misregistration of Virtual Human Audio: Implications of the Precedence Effect

David M. Krum*
USC Institute for Creative Technologies

Evan A. Suma†
USC Institute for Creative Technologies

Mark Bolas‡
USC School for Cinematic Arts
USC Institute for Creative Technologies

Abstract

Virtual humans are often presented as mixed reality characters projected onto screens that are blended into a physical setting. Stereo loudspeakers to the left and right of the screen are typically used for virtual human audio. Unfortunately, stereo pairs can produce an effect known as precedence, which causes users standing close to a particular loudspeaker to perceive a collapse of the stereo sound to that singular loudspeaker. We studied if this effect might degrade the presentation of a virtual character, or if this would be prevented by the ventriloquism effect. Our results demonstrate that from viewing distances common to virtual human scenarios, a movement equivalent to a single stride can induce a stereo collapse, creating conflicting perceived locations of the virtual human's voice. Users also expressed a preference for a sound source collocated with the virtual human's mouth rather than a stereo pair. These results provide several design implications for virtual human display systems.

Index Terms: H.5.1 [Information Systems]: Multimedia Information Systems—Artificial, augmented, and virtual realities; B.4.2 [Input/Output and Data Communications]: Input/Output Devices—Voice.

1 Introduction

Computer controlled virtual humans are an increasingly important component of applications such as training, therapy, novel computer interfaces, and social research. Often a 3D mixed reality presentation is preferred, where life-sized virtual characters are blended into a physical setting. Digital projectors are often used, due to their practicality and low cost.

While a single loudspeaker located near a character's mouth can be used to portray the voice of a single virtual human, this placement can be problematic. With a rear projected screen, the loudspeaker would block the video image. Furthermore, placing loudspeakers behind the screen would muffle the audio. While there are perforated screens that allow sounds to pass through, these screens require front projection, unfortunately allowing users to cast shadows across the character's image. A rear projection display combined with a stereo loudspeaker pair is thus a common configuration for virtual human installations.

As with any stereo pair, this configuration is subject to the precedence effect [9]. The precedence effect occurs when a listener stands much closer to one of the two loudspeakers in a stereo pair. The wavefront from the nearby loudspeaker will arrive sooner than, or precedes, the other loudspeaker's wavefront. The human perceptual system has echo cancellation, causing the second wavefront to be ignored. Only the initial wavefront is perceived, causing the perceived sound location to collapse to the nearby loudspeaker, breaking the stereo spatialization. We were concerned that the precedence effect might cause a virtual human's voice to shift left and right as the listener moves around. Adding to our uncertainty was the ventriloquism effect, which might counteract the precedence effect. The ventriloquism effect can create the perception that a voice or sound, generated elsewhere, is emanating from the visual image of a temporally related source [3, 5].

It is certainly unrealistic if a character's voice emanates from a point that is offset from the character's mouth. This offset might create a cognitive dissonance which might weaken the illusion of realism, negatively affect conversational interactions, and cause participants to disengage. Our aim was to examine the impact of precedence effect in a mixed reality virtual human presentation and determine the design implications of these findings.

2 Related Work

Background work in spatial audio reproduction includes headphone based techniques using binaural audio and head related transfer functions [4] as well as techniques using arrays of loudspeakers like Ambisonics [8] and wavefield synthesis [2]. Headphone based techniques are less appealing since they require users to wear an additional device, interfering with simple "walk up and interact" experiences. Wavefield synthesis requires large numbers of loudspeakers, perhaps hundreds or more, increasing cost and complexity. Ambisonics typically require four or more loudspeakers, as well as decoding hardware, and can have reproduction issues in large spaces without sound treatment to control echo and reverberation. Additional related work on virtual human audio examines temporal synchronization [7] and high quality visual and audio systems [6].

3 Methods

To determine if the precedence effect can alter a user's perception of a projected virtual human, we designed and conducted a mixed design study where participants listened to a virtual human reading literary passages. Participant were divided into three groups, with three physical locations in front of the virtual human representing the between subjects condition. Audio presentation (left, center, or stereo) was the within subjects condition.

In the first phase of the experiment, participants listened to three short passages from the virtual character, presented randomly on the center, left, or stereo loudspeakers. In the second phase, participants listened and compared pairs of sentences, alternatively presented by the center loudspeaker and a stereo pair in random order.

Thirty-six participants, over the age of 18, with 20/20 corrected vision and hearing in both ears were recruited through email and the Craigslist website. The gender ratio was evenly balanced (18 male and 18 female), and participant age ranged from 20 years to 67 years, with a mean of 37.6 and a standard deviation of 13.6.

3.1 Apparatus

Participants were placed at one of three positions approximately 12 feet in front of the virtual character (see Figure 1). A three foot lateral spacing was chosen to represent the distance of a single stride and a comfortable interpersonal distance between individuals.

The virtual human in the study was a male soldier using a Cepstral text-to-speech voice and rendered by the Panda3D engine.

*e-mail: krum@ict.usc.edu
†e-mail: suma@ict.usc.edu
‡e-mail: bolas@ict.usc.edu

IEEE Symposium on 3D User Interfaces 2012
4-5 March, Orange County, CA, USA
978-1-4673-1205-9/12/$31.00 ©2012 IEEE

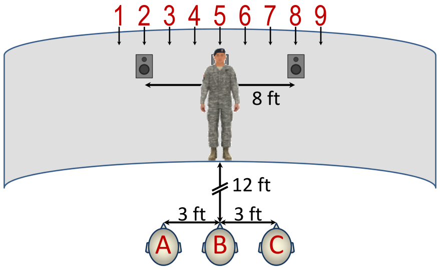

Figure 1: Three loudspeakers located behind a sound transparent screen provided audio to the left of the virtual human, at the center, or in stereo. Participants stood at A, B, or C.

Passages were sampled from *Moby Dick*. The character was projected onto a perforated curved screen (see Figure 1), which allowed sounds to pass through. Three Mackie HR824 studio monitor loudspeakers were placed behind the screen at the center, 4 feet to the left, and 4 feet to the right. Sound pressure levels were calibrated to match between audio conditions. Some sound treatment was applied behind the loudspeakers and screen to limit reverberations.

3.2 Procedure

In the first phase of the study, each participant listened to three short passages read by the character and presented by either the left, center, or stereo loudspeakers. The order of the audio presentation was fully randomized. After each passage, participants were asked two co-presenece questions, based on the Bailenson et al. social presence questionaire [1]. Participants then indicated the apparent horizontal location of the voice by referencing nine numbers placed along the top of the screen. The numbers were spaced approximately 16 inches apart, with the number 5 located at the center, directly above center loudspeaker and the visual image of the character. Numbers 2 and 8 indicated the left and right loudspeakers.

In the second phase, participants listened to the character's delivery of four sentence pairs. Each sentence pair consisted of the same sentence, repeated twice, and alternating in random order between stereo and center loundspeakers. Participants were asked, for each pair, "Which line was delivered more like a real person?"

4 Results and Discussion

A mixed ANOVA statistical test was performed to determine if the within-subjects condition of audio presentation as well as the between-subjects condition of listener position created significant differences in the perceived location of the sound source at the $\alpha = .05$ level. Since Mauchly's Test of Sphericity indicated a possible violation of sphericity for the within-subject effects of audio presentation, we performed a Greenhouse-Geisser correction.

A significant main effect of audio presentation (left, center, or stereo) was observed in perceived location, $F(2,66) = 32.40, p < .001, \eta_p^2 = .50$. Examining the 95% confidence intervals reveals that the perceived location of speech produced by the left loudspeaker, 95% CI [3.27, 4.45], is well separated and clearly different from the center [5.92, 6.35] and stereo [5.24, 6.09] presentations.

The between subjects variable, listener position (A:Left, B:Middle, or C:Right) was observed to create a significant difference in perceived location $F(2,33) = 5.56, p = .008, \eta_p^2 = .25$. As visible in the crossing of the trendlines (see Figure 2), a significant interaction effect can be seen in the data for the stereo condition in combination with listener position $F(4,66) = 10.16, p < .001, \eta_p^2 = .38$. Listeners at the rightmost or leftmost positions perceived a shift of the stereo sound source towards the same corresponding right or left sides of the screen. This is evidence of the precedence effect.

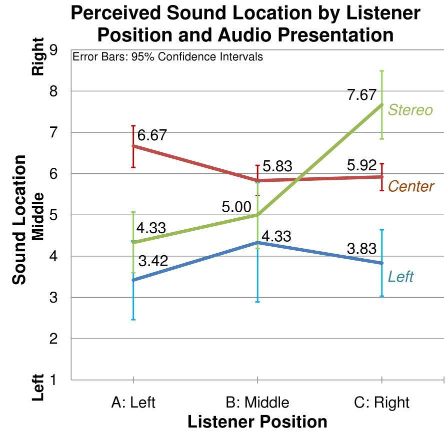

Figure 2: A significant interaction in perceived location was observed for listener's position and stereo. Listeners at the left localized the stereo to the left. Listeners at the right localized it to the right.

In the second phase of the study, the sentence pair trials, participants expressed a preference for the center loudspeaker over the stereo pair, regardless of listener position. The center loudspeaker was chosen over stereo for realism 103 times out of 144 trials (71.5%). This trend was also replicated within participants grouped by listener position (Left: 87.5%, Middle: 64.6%, Right: 62.5%). This suggests that a single loudspeaker delivered better realism for the voice, improving the character's co-presence.

5 Conclusion and Future Work

This poster demonstrates that the precedence effect can occur in a typical virtual human installation with stereo audio, causing misperception of the audio source. The ventriloquism effect does not appear to mask this misperception to any major extent. Accordingly, designers should examine the range of motion and number of users required and select designs that can robustly collocate the virtual human audio and visuals. Perforated screens and individual loudspeakers for each virtual character should be considered.

References

[1] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence-Teleop. Virt.*, 10:583–598, Dec 2001.

[2] A. J. Berkhout. A holographic approach to acoustic control. *J. Audio Eng. Soc*, 36(12):977–995, 1988.

[3] P. Bertelson. Chapter 14 ventriloquism: A case of crossmodal perceptual grouping. In T. B. Gisa Aschersleben and J. Msseler, editors, *Cognitive Contributions to the Perception of Spatial and Temporal Events*, volume 129 of *Advances in Psychology*, pages 347 – 362. North-Holland, 1999.

[4] J. Blauert. *Räumliches Hören (Spatial Hearing)*. S. Hirzel-Verlag, Stuttgart, Germany, 1974.

[5] C. Choe, R. Welch, R. Gilford, and J. Juola. The ventriloquist effect: Visual dominance or response bias? *Atten. Percept. Psycho.*, 18:55–60, 1975. 10.3758/BF03199367.

[6] M. Courgeon, M. Rebillat, B. Katz, C. Clavel, and J.-C. Martin. Life-sized audiovisual spatial social scenes with multiple characters: Marc & smart-i2. In *Meeting of the French Association for Virtual Reality*, 2010.

[7] C. Ennis, R. McDonnell, and C. O'Sullivan. Seeing is believing: body motion dominates in multisensory conversations. *ACM T. Graphic.*, 29:91:1–91:9, July 2010.

[8] P. Fellget. Ambisonics. part one: General system description. *Studio Sound*, 17:20–22,40, August 1975.

[9] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman. The precedence effect. *J. Acoust. Soc. Am.*, 106:1633–1654, 1999.