Movements and voices affect perceived sex of virtual conversers

Rachel McDonnell* Carol O'Sullivan†

Graphics, Vision and Visualisation Group, Trinity College Dublin.

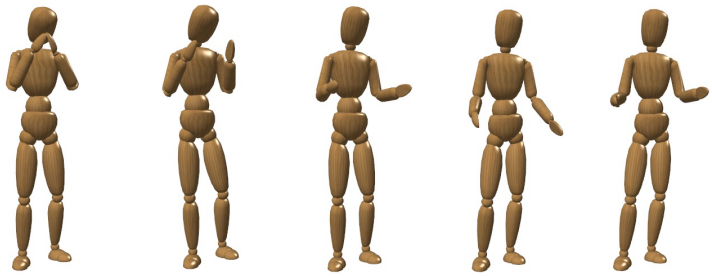

Figure 1: *Androgynous mannequin used in experiments.*

Abstract

In this paper, we investigate the ability of humans to determine the sex of conversing characters, based on audio and visual cues. We used a corpus of motions and sounds captured from three male and three female actors conversing about a range of topics. In our Unisensory Experiments, visual and auditory stimuli were presented separately to participants who rated how male or female they found them to be. In our Multisensory Experiments, audio and visual information were integrated to determine how they interacted. We found that audio was much easier to classify than motion, and that audio affected but did not saturate ratings when motion and audio were integrated. Finally, even when informative appearance cues were present, this did not help to disambiguate incongruent motion and audio.

CR Categories: I.3.7 [Computer Graphics]: Three Dimensional Graphics and Realism—Animation;

Keywords: perception, motion capture, virtual humans

1 Introduction

Due to the increase in popularity of interactive drama video games (such as Heavy Rain or L.A. Noire), delivering plausible conversing virtual characters has now become extremely important. In some large budget games, performance actors provide the voices, movements and even appearance of the virtual character they are driving, but this is not always possible due to budget, time or location constraints. One important factor to consider when mismatching voices and motions is that people find it disturbing if there are discrepancies between how a character appears and how they act [Vinayagamoorthy et al. 2006]. Furthermore, conflicting auditory and visual cues may affect the perception of gender of the character [van der Zwan et al. 2009]. To our knowledge, it has not been established if other motions besides walking can be classified as male or female based on body motion alone. In this paper, we conduct a series of perceptual experiments in order to investigate some of the factors that affect the perceived sex of virtual conversing characters.

In previous work [McDonnell et al. 2007], we showed that for walking motions, the sex of the motion captured actor needed to match that of the model it was applied to, or else the viewer judged the resulting stimulus as ambiguous. We also found that for an androgynous character, motion tended to dominate. In that study, we tested only the interaction between body motion and appearance for walking gaits, with no audio.

2 Background

There has been previous research in the field of experimental psychology on determining the sex of a human representation, based on body motion alone. Kozlowski and Cutting [1977] showed that natural motion is a sufficient cue to determine the sex of a walking human, using "point light" displays. These displays consisted of twelve moving light sources only so no form information was present. Since then, it has been shown that point light displays contain enough information to recognise a *particular* individual [Cutting and Kozlowski 1977] or even one's own walking pattern [Beardsworth and Buckner 1981]. The influence of body shape on the perception of sex has also been investigated. Mather and Murdoch [1994] tested sex perception on a range of point light displays with differing torso shapes, and found that motion was a more salient cue than shape. Johnson and Tassinary [2005] found the opposite to be true, which may have been due to the fact that their stimuli had stronger shape cues.

There has also been some work in the literature on perception of conversing humans. Briton and Hall [1995] found evidence to suggest that there are differences between the gestures of men and women when conversing. Also, it has been found that talker and listener roles can be identified from biological motion alone [Rose and Clarke 2009]. However, to our knowledge there have been no previous attempts to determine if the sex of a converser can be identified by motion alone.

Most related to our work are the multisensory experiments performed by van der Zwan et al. [2009]. They used auditory stimuli of actual footsteps from a range of females and males wearing different footwear. Their visual stimuli depicted a range of male and female point light walkers. They found that auditory cues can influence perceived gender in biological motion displays, particularly when the gender of the visual stimuli were ambiguous. We wish to determine if the same is true for conversing motions and sounds.

*Rachel.McDonnell@cs.tcd.ie

†Carol.OSullivan@cs.tcd.ie

APGV 2010, Los Angeles, California, July 23 – 24, 2010.
© 2010 ACM 978-1-4503-0248-7/10/0007 $10.00

3 Virtual Stimuli

We used data previously captured from two sets of actors; the first set consisted of three males and the second had three females. As in [McDonnell et al. 2009], actors stood around the corners of a triangle and conversations were free-flowing and not scripted. All actors were non-professionals and were familiar with the motion capture setup and environment. An omni-directional microphone placed in the centre of the triangle recorded the audio. We used five conversations with dominant speakers (i.e., where only one person talked and the others politely listened) of one minute length each, per actor. We chose a model that was judged as androgynous in previous studies [McDonnell et al. 2007] to display our motion (Figure 1), as it was the motion and not the form that we were interested in testing.

The model was viewed on a white background, for good contrast. We used a frontal view since Halevina and Troje [2007] found that it facilitates sex classification. The experimental system was developed using an open-source renderer and a commercially available animation engine. The experiment was run on a workstation with 2GB of RAM, an 8-series GeForce graphics card on a wide-screen 24-inch LCD monitor.

4 Unisensory Experiments

Our first experiment was a baseline test which was conducted to investigate sex perception of motion and audio independently. The purpose was to collect classification data for a large range of motion and audio clips, so that we could choose an appropriate subset for our Multisensory Experiments (Section 5).

In the Motion Baseline, we wished to determine if it is possible to identify the sex of a person based on conversational body motion alone. In the Audio Baseline, we investigated the ability of participants to rate the sex of the different voices. We hypothesised that audio would be easier to rate and that there would be a correlation between ratings for the voice and audio of the same actor.

Twelve participants (4F, 8M) took part in this experiment. As in all succeeding experiments, participants were naïve to the purpose of the experiment, from different educational backgrounds, and were given book vouchers as compensation. Each participant viewed the motion and audio baseline blocks separately. We counterbalanced the order of the blocks to avoid ordering effects.

4.1 Motion Baseline

Ten motion clips were chosen from each actor (two from each of the five conversations) and applied to the androgynous figure. We ensured that each of the clips contained motion of the actor talking, not listening. Participants viewed, in random order, 120 trials in total (10 motion clips * 6 actors * 2 repetitions). Each trial was displayed for five seconds after which participants were asked to categorise the motion they just saw on a 5-point scale: 1: *very male*, 2: *male*, 3: *ambiguous*, 4: *female*, 5: *very female*. The next trial was shown once a key-press was recorded. Participants were told in advance that the appearance of the character was neutral and to therefore base their judgements on the motion. A fixation cross was displayed in the centre of the screen before every trial.

4.2 Audio Baseline

Using auditory information of foot-steps alone, Li et al. [1991] showed that gender was correctly identified 72% of the time. In our Audio Baseline, we wished to determine how different audio clips of actors' voices would be classified on a 5-point rating scale. Ten audio clips were chosen from each actor (two from each of the five conversations). We ensured that the audio clips contained the voice of only that specific actor while talking. Participants were given a set of Sennheiser HD 202 headphones and were asked to listen to, in random order, 120 trials in total (10 audio clips * 6 voices * 2 repetitions). A white screen was displayed while they listened to the audio, so as not to influence their decisions with any visual information. Trials were five seconds long, and participants rated the audio on the same 5-point scale as before. They were told in advance that there would be no visual information and to base their judgements on the tone of voice in the audio clip.

4.3 Results

In order to first determine if participants could in fact distinguish between male and female motions, we averaged ratings over all repetitions of female and male motions, for each participant. A repeated measures ANOVA showed a main effect of *motion type* (male or female) ($F_{1,11} = 60, p < 0.00001$). This demonstrates that, based on body motion alone, participants were able to distinguish between male and female conversational body motions.

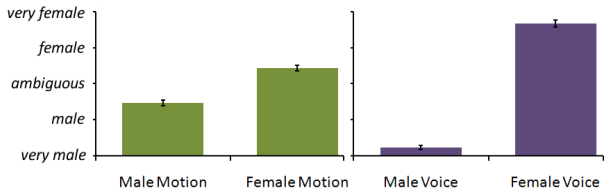

Figure 2: *Main effect of (L) motion type, and (R) voice type.*

Using the same procedure, we also found a strong main effect of *voice type* (male or female) ($F_{1,11} = 578, p < 0.00001$). Figure 2 shows that voices were easier to classify than motions. There was no interaction between motion and voice type.

4.3.1 Actors

Next, we wished to determine whether the six actors were rated differently. We hypothesised that there would be differences in ratings for both the audio and motion, and that there would be a correlation between ratings for the voices and motions of each actor.

A two-way repeated measures ANOVA with conditions *modality* (voice or motion) and *actor* (6) showed a main effect of actor ($F_{5,55} = 272, p < 0.00001$). Post-hoc analysis for this and all subsequent tests was conducted using Newman-Keuls pairwise comparisons of means. Overall, Male Actor 1 (MA1) was rated as the most male actor, and MA2 and MA3 were equally rated as less male. Female Actor 1 (FA1) was rated as the most female, FA3 next, and FA2 as the least female ($p < 0.02$ in all cases).

We found an interaction between modality and actor ($F_{5,55} = 165, p < 0.00001$). Out of the female actors there were three significantly different ratings for motion (FA1 was the most female, FA3 next and FA2 the least female, $p < 0.0002$ in all cases), but all voices were rated as equally female. Out of the males, there were three significantly different ratings for the motions (MA1 was the most male, then MA3, and MA2 was the least male, $p < 0.03$ in all cases). However, MA1 and MA2 were rated equally male for the voices, with MA3 being significantly less male ($p < 0.02$). Therefore, there did not appear to be any correlation between the ratings for a particular actor's voice and motion (Figure 3).

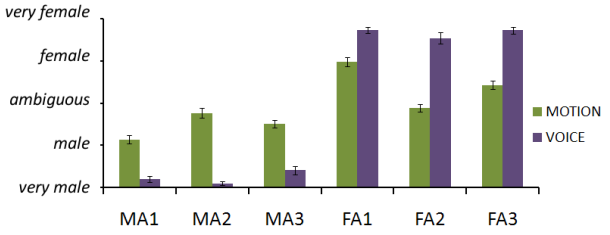

Figure 3: *Average ratings for the 10 motion and audio clips of the three male (MA1 - MA3) and female actors (FA1 - FA3).*

4.3.2 Motion Categorisation

In order to choose a set of motion clips for the Multisensory Experiments, we averaged the ratings of each of the 10 clips per actor, over all repetitions. We sampled the set into five categories: **most male**, **male**, **ambiguous**, **female**, and **most female**; choosing the three most appropriately rated clips per category. We will refer to this subset of 15 motion clips as the *categorised motions*.

5 Multisensory Experiments

In this set of experiments, we wished to examine the perception of the sex of more complex stimuli, incorporating both auditory and visual information. The first experiment investigates how combining audio with motion will effect sex perception of a character with non-informative shape information (Figure 1). The second experiment explores this further, but with more informative shape information (Figure 5). In particular, we wished to determine if the perception of the motion could be altered or inverted, by presenting participants with both audio and appearance cues.

We hypothesised that: audio would have the strongest effect on sex perception and that this would reduce the sensitivity to body motion differences; furthermore, when male and female appearance cues were added, motion cues would become even less salient.

5.1 Motion and Audio

We used only the categorised motions in this experiment. Since there were only small differences between ratings for the actors' audio clips, we randomly chose a female or a male voice to pair with each of the motions. Eighteen new volunteers that had not taken part in the previous experiment took part in this experiment (7F, 11M). Each participant simultaneously viewed and listened to 60 trials in random order: 15 motions (5 categories * 3 clips) * 2 audio types (male and female) * 2 repetitions. Participants were told that the appearance of the character would be neutral, and to therefore take the motion and tone of the audio into account when making their decisions. As before, a 5-point scale ranging from *very male* to *very female* was used.

5.1.1 Results

Of most interest to us was whether or not the addition of audio affected the ratings of the categorised motions so we compared the multisensory with the unisensory results. We conducted two between-groups ANOVAs, where data for the *no audio* condition was taken from the results of the Motion Baseline.

First, we tested the effect of adding male audio on the perceived sex of the character. We averaged the results for each repetition of each of the categorised motions with *no audio* and with *male audio*. A two-way ANOVA with between-subjects condition *no/male audio* (2) and within subjects condition *motion category* (5) was conducted. A main effect of motion category ($F_{4,112} = 147, p < 0.00001$) was found, where overall, each of the motions was categorised significantly differently ($p < 0.0002$ in all cases). This is as expected since we purposely chose previously categorised motions that were significantly different in sex rating from each other.

A main effect of male audio was also found ($F_{1,28} = 23, p < 0.00007$), where overall motions were rated as more male when the male audio was present, than when no audio was present. An interaction also occurred ($F_{4,112} = 8, p < 0.00002$), which was due to a plateau effect where **most male** was rated equally to **male** when male audio was present (see Figure 4, blue series).

Next, we tested the effect of adding female audio. As before, a main effect of motion category was found ($F_{4,112} = 127, p < 0.00001$) where all motions were rated differently. A main effect of female audio was also found, with the addition of female audio making the ratings more female overall ($F_{1,28} = 34, p < 0.00001$). Finally, an interaction occurred ($F_{4,112} = 10, p < 0.00001$) which was again due to a plateau effect where **most female** was rated equally to **female** when female audio was present (see Figure 4, red series)

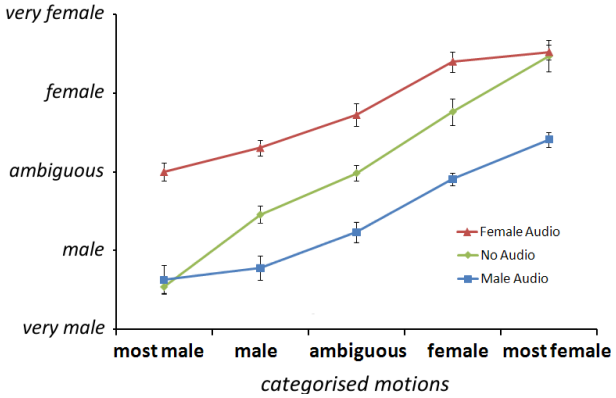

Figure 4: *Average ratings for motion clips with male audio (blue series), no audio (green series) and female audio (red series).*

Our results indicate that audio affects the perception of sex, except when paired with congruent motion that has already strong cues to indicate sex. Male audio was never sufficient to convince participants that a character with female motion was male. Equally, female audio was never sufficient to convince participants that a character with male motion was in fact a female. Both cases were judged as ambiguous. However, when viewing ambiguous motions, participants did judge them to be more male and female with audio cues (as in [van der Zwan et al. 2009]). This also mirrors our previous work where appearance dominated for the synthetic neutral motions [McDonnell et al. 2007].

However, a floor or ceiling effect did not occur, so it is clear that the audio was not the only discriminating factor. Our second Multisensory Experiment was conducted in order to determine if adding appearance cues would shift ratings of incongruent motion and audio away from an "ambiguous" rating.

5.2 Appearance, Motion and Audio

In [McDonnell et al. 2007], we found that the appearance of the character affected sex perception. Therefore, in this experiment, we tested if displaying the categorised motions and audio on a character with clear male or female appearance (Figure 5) would alter the ratings. Audio and appearance were kept congruent throughout this experiment, so we hypothesised that with incongruent motion ap-

Figure 5: *Man and woman characters used in the Appearance, Motion and Audio Experiment.*

plied, participants might be more inclined to rate the character in line with the audio and appearance cues.

As before, we used the categorised motions, and male and female audio. In this experiment, male voices were paired with a man character and female voices with a woman character, rather than the androgynous figure. A male voice was never paired with the woman and vice versa. Ten volunteers (4F, 6M) took part in this experiment, and were asked as before to rate the stimuli on a 5-point scale from *very male* to *very female*. Each participant simultaneously viewed and listened to 60 trials in random order: 15 motions (5 categories * 3 clips) * 2 audio-body types (male audio with a man character, and female audio with a woman character) * 2 repetitions. This time they were told to take all information into account, including appearance, audio and motion.

5.2.1 Results

A two-way ANOVA with between-subjects condition *body type* (androgynous character or man) and within subjects condition *motion category* (5) was conducted on all data with male audio. A main effect of motion category was found, as before ($F_{4,104} = 80, p < 0.00001$). No effect of body type was found, nor was there an interaction between body type and motion category. A second two-way ANOVA was conducted on the data with female audio. Again, a main effect of motion was found ($F_{4,104} = 54, p < 0.00001$). No effect of body type was found, nor any interactions. This implies that the addition of appearance cues did not help to disambiguate incongruent motion and audio. Also, surprisingly, the ambiguous motions did not appear any more male or female with combined audio and appearance, than they did with audio alone.

6 Discussion and Future Work

In order to create unambiguous, plausible conversing characters, we found that voice affects sex perception, but does not dominate. An ambiguous motion can be disambiguated with the addition of audio, but incongruent audio and motion will always be considered as ambiguous. The addition of extra appearance cues does not affect ratings. These results will be useful for computer graphics applications, since it is now clear that conversational motions carry sex information and should only be displayed with a congruent voice to avoid ambiguity. Also, pairing a voice with a motion is an effective way to increase how male/female a character appears.

In this paper we did not test the affect of incongruent audio and appearance, or appearance and motion without audio, but these could be interesting conditions to test in the future. The affect of facial and hand motion will also be investigated in future work.

In post-experiment questioning, some participants reported that they used the pose of the character and the wrists to determine if the motion was male or female. Others mentioned cues such as the speed of the gestures or how far apart the feet were. In order to create realistic synthetic conversing characters, it will be important to find a robust classification for our set of motions (such as [Troje 2002]).

References

BEARDSWORTH, T., AND BUCKNER, T. 1981. The ability to recognize oneself from a video recording of ones movements without seeing ones body. *Bulletin of the Psychonomic Society 18*, 1, 19–22.

BRITON, N., AND HALL, J. 1995. Beliefs about female and male nonverbal communication. *Sex Roles: A Journal of Research 32*, 1–2, 79–90.

CUTTING, J., AND KOZLOWSKI, L. 1977. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society 9*, 5, 353–356.

HALEVINA, A., AND TROJE, N. 2007. Sex-classification of point-light walkers: Viewpoint, structure, kinematics. In *Poster presented at Vision Science Society meeting, Sarasota, FL.*

JOHNSON, K. L., AND TASSINARY, L. G. 2005. Perceiving sex directly and indirectly: Meaning in motion and morphology. *Psychological Science 16*, 11, 890–897.

KOZLOWSKI, L., AND CUTTING, J. 1977. Recognizing the sex of a walker from a dynamic point-light display. *Perception and Psychophysics 21*, 6, 578–580.

LI, X., LOGAN, R. J., AND PASTORE, R. E. 1991. Perception of acoustic source characteristics: walking sounds. *The Journal of the Acoustical Society of America 90*, 6, 3036–3049.

MATHER, G., AND MURDOCH, L. 1994. Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London, Series B: Biological Sciences 258*, 1358, 273–279.

MCDONNELL, R., JÖRG, S., HODGINS, J. K., NEWELL, F., AND O'SULLIVAN, C. 2007. Virtual shapers & movers: form and motion affect sex perception. In *APGV '07: Proceedings of the 4th symposium on Applied perception in graphics and visualization*, 7–10.

MCDONNELL, R., ENNIS, C., DOBBYN, S., AND O'SULLIVAN, C. 2009. Talking bodies: Sensitivity to de-synchronization of conversations. *ACM Transactions on Applied Perception 22*, 1, 22:1–22:8.

ROSE, D., AND CLARKE, T. J. 2009. Look who's talking: Visual detection of speech from whole-body biological motion cues during emotive interpersonal conversation. *Perception 38*, 1, 153–156.

TROJE, N. F. 2002. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision 2*, 5, 371–387.

VAN DER ZWAN, R., MACHATCH, C., KOZLOWSKI, D., TROJE, N. F., BLANKE, O., AND BROOKS, A. 2009. Gender bending: auditory cues affect visual judgements of gender in biological motion displays. *Experimental Brain Research 198*, 2-3, 373–382.

VINAYAGAMOORTHY, V., GILLIES, M., STEED, A., TANGUY, E., PAN, X., LOSCOS, C., AND SLATER, M. 2006. Building expression into virtual characters. *Eurographics State of the Art Reports*.