Feel the beat: using cross-modal rhythm to integrate perception of objects, others, and self

Paul Fitzpatrick **Artur Arsenio**
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

Abstract

For a robot to be capable of development, it must be able to explore its environment and learn from its experiences. It must find (or create) opportunities to experience the unfamiliar in ways that reveal properties valid beyond the immediate context. In this paper, we develop a novel method for using the rhythm of everyday actions as a basis for identifying the characteristic appearance and sounds associated with objects, people, and the robot itself. Our approach is to identify and segment groups of signals in individual modalities (sight, hearing, and proprioception) based on their rhythmic variation, then to identify and bind causally-related groups of signals across different modalities. By including proprioception as a modality, this cross-modal binding method applies to the robot itself, and we report a series of experiments in which the robot learns about the characteristics of its own body.

1. Introduction

To robots and young infants, the world is a puzzling place, a confusion of sights and sounds. But buried in the noise there are hints of regularity. Some of this is natural; for example, objects tend to go *thud* when they fall over and hit the ground. Some is due to the child; for example, if it shakes its limbs in joy or distress, and one of them happens to pass in front of its face, it will see a fleshy blob moving in a familiar rhythm. And some of the regularity is due to the efforts of a caregiver; consider an infant's mother trying to help her child learn and develop, perhaps by tapping a toy or a part of the child's body (such as its hand) while speaking its name, or making a toy's characteristic sound (such as the *bang-bang* of a hammer).

In this paper we seek to extract useful information from repeated actions performed either by a caregiver or the robot itself. Observation of infants shows that such actions happen frequently, and from a computational perspective they are ideal learning material since they are easy to identify and offer a wealth of redundancy (important for robustness). The information we seek from repeated actions are the characteristic appearances and sounds of the object, person, or robot involved, with context-dependent information such as the visual background or unrelated sounds stripped away. This allows the robot to generalize its experience beyond its immediate context and, for example, later recognize the same object used in a different way.

We wish our system to be scalable, so that it can correlate and integrate multiple sensor modalities (currently sight, sound, and proprioception). To that end, we detect and cluster periodic signals within their individual modalities, and only then look for cross-modal relationships between such signals. This avoids a combinatorial explosion of comparisons, and means our system can be gracefully extended to deal with new sensor modalities in future (touch, smell, etc).

This paper begins by introducing our robotic platform and what it can sense. We then introduce the methods we use for detecting regularity in individual modalities and the tests applied to determine when to 'bind' features in different modalities together. The remainder (and larger part) of the paper presents experiments where the robot detects regularity in objects, people it encounters, and finally itself.

2. Platform and percepts

This work is implemented on the humanoid robot Cog (Brooks et al., 1999). Cog has an active vision head, two six-degree of freedom arms, a rotating torso, and a microphone array arranged along its shoulders. For this paper, we work with visual input from one of Cog's four cameras, acoustic input from the microphone array, and proprioceptive feedback from joints in the head, torso, and arms.

Figure 1 shows how the robot's perceptual state can be summarized – the icons shown here will be used throughout the paper. The robot can detect

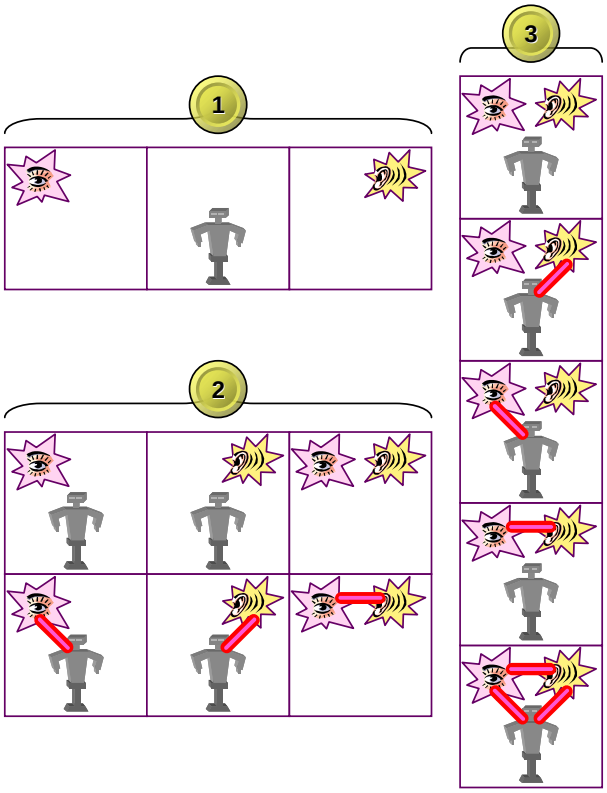

Figure 1: A summary of the possible perceptual states of our robot – the representation shown here will be used throughout the paper. Events in any one of the three modalities (sight, proprioception, or hearing) are indicated as in block **1**. When two events occur in different modalities, they may be independent (top of **2**) or bound (bottom of **2**). When events occur in three modalities, the possibilities are as shown in **3**.

periodic events in any of the individual modalities (sight, hearing, proprioception). Any two events that occur in different modalities will be compared, and may be grouped together if there is evidence that they are causally related or *bound*. Such relations are transitive: if events A and B are bound to each other, and B and C are bound to each other, then A and C will also be bound. This is important for consistent, unified perception of events.

This kind of summarization ignores cases in which there are, for example, multiple visible objects moving periodically making different sounds. We return to this point later in the paper. We have previously demonstrated that our system can deal well with multiple-binding cases, since it performs segmentation in the individual modalities (Arsenio and Fitzpatrick, 2003). For this paper, there is no real need to consider such cases, since we don't expect the robot's caregiver to maliciously introduce distractors into its environment – but nevertheless it is an important feature of our algorithm, which we now present.

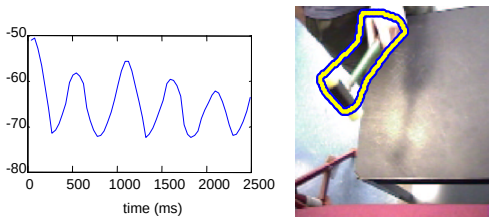

Figure 2: When watching a person using a hammer, the robot detects and group points moving in the image with similar periodicity (Arsenio et al., 2003) to find the overall trajectory of the hammer and separate it out from the background. The detected trajectory is shown on the left (for clarity, just the coordinate in the direction of maximum variation is plotted), and the detected object boundary is overlaid on the image on the right.

3. Detecting periodic events

We are interested in detecting conditions that repeat with some roughly constant rate, where that rate is consistent with what a human can easily produce and perceive. This is not a very well defined range, but we will consider anything above 10Hz to be too fast, and anything below 0.1Hz to be too slow. Repetitive signals in this range are considered to be *events* in our system. For example, waving a flag is an event, clapping is an event, walking is an event, but the vibration of a violin string is not an event (too fast), and neither is the daily rise and fall of the sun (too slow). Such a restriction is related to the idea of natural kinds (Hendriks-Jansen, 1996), where perception is based on the physical dimensions and practical interests of the observer.

To find periodicity in signals, the most obvious approach is to use some version of the Fourier transform. And indeed our experience is that use of the Short-Time Fourier Transform (STFT) demonstrates good performance when applied to the visual trajectory of periodically moving objects (Arsenio et al., 2003). For example, Figure 2 shows a hammer segmented visually by tracking and grouping periodically moving points. However, our experience also leads us to believe that this approach is not ideal for detecting periodicity of *acoustic* signals. Of course, acoustic signals have a rich structure around and above the kHz range, for which the Fourier transform and related transforms are very useful. But detecting gross repetition around the single Hz range is very different. The sound generated by a moving object can be quite complicated, since any constraints due to inertia or continuity are much weaker than for the physical trajectory of a mass moving through space. In our experiments, we find that acoustic signals may vary considerably in amplitude between repetitions, and that there is significant variability or drift in the length of the pe-

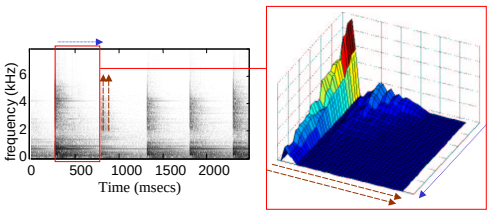

Figure 3: Extraction of an acoustic pattern from a periodic sound (a hammer banging). The algorithm for signal segmentation is applied to each normalized frequency band. The box on the right shows one complete segmented period of the signal. Time and frequency axes are labeled with single and double arrows respectively.

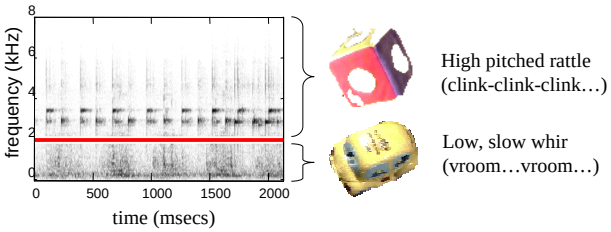

Figure 4: Results of an experiment in which the robot could see a car and a cube, and both objects were moving – the car was being pushed back and forth on a table, while the cube was being shaken (it has a rattle inside). By comparing periodicity information, the high-pitched rattle sound and the low-pitched *vroom* sound were distinguished and bound to the appropriate object, as shown on the spectrogram. The object segmentations shown were automatically determined.

riods. These two properties combine to reduce the efficacy of Fourier analysis. This led us to the development of a more robust method for periodicity detection, which is now described. In the following discussion, the term *signal* refers to some sensor reading or derived measurement, as described at the end of this section. The term *period* is used strictly to describe event-scale repetition (in the Hz range), as opposed to acoustic-scale oscillation (in the kHz range).

Period estimation – For every sample of the signal, we determine how long it takes for the signal to return to the same value from the same direction (increasing or decreasing), if it ever does. For this comparison, signal values are quantizing adaptively into discrete ranges. Intervals are computed in one pass using a look-up table that, as we scan through the signal, stores the time of the last occurrence of a value/direction pair. The next step is to find the most common interval using a histogram (which requires quantization of interval values), giving us an initial estimate $p_{estimate}$ for the event period. This is essentially the approach presented in (Arsenio and Fitzpatrick, 2003). For the work presented in this paper, we extended this method to explicitly take into account the possibility of drift and variability in the period, as follows.

Clustering – The previous procedure gives us an estimate $p_{estimate}$ of the event period. We now cluster samples in rising and falling intervals of the signal, using that estimate to limit the width of our clusters but not to constrain the distance between clusters. This is a good match with real signals we see that are generated from human action, where the periodicity is rarely very precise. Clustering is performed individually for each of the quantized ranges and directions (increasing or decreasing), and then combined afterwards. Starting from the first signal sample not assigned to a cluster, our algorithm runs iteratively until all samples are assigned, creating new clusters as necessary. A signal sample extracted at time t is assigned to a cluster with center c_i if $\| c_i - t \|_2 < p_{estimate}/2$. The cluster center is the average time coordinate of the samples assigned to it, weighted according to their values.

Merging – Clusters from different quantized ranges and directions are merged into a single cluster if $\| c_i - c_j \|_2 < p_{estimate}/2$ where c_i and c_j are the cluster centers.

Segmentation – We find the average interval between neighboring cluster centers for positive and negative derivatives, and break the signal into discrete periods based on these centers. Notice that we do not rely on an assumption of a *constant* period for segmenting the signal into repeating units. The average interval is the final estimate of the signal period.

The output of this entire process is an estimate of the period of the signal, a segmentation of the signal into repeating units, and a confidence value that reflects how periodic the signal really is. The period estimation process is applied at multiple temporal scales. If a strong periodicity is not found at the default time scale, the time window is split in two and the procedure is repeated for each half. This constitutes a flexible compromise between both the time and frequency based views of a signal: a particular movement might not appear periodic when viewed over a long time interval, but may appear as such at a finer scale.

Figure 2 shows an example of using periodicity to visual segment a hammer as a human demonstrates the periodic task of hammering, while Figure 3 shows segmentation of the sound of the hammer in the time-domain. Segmentation in the frequency-domain was demonstrated

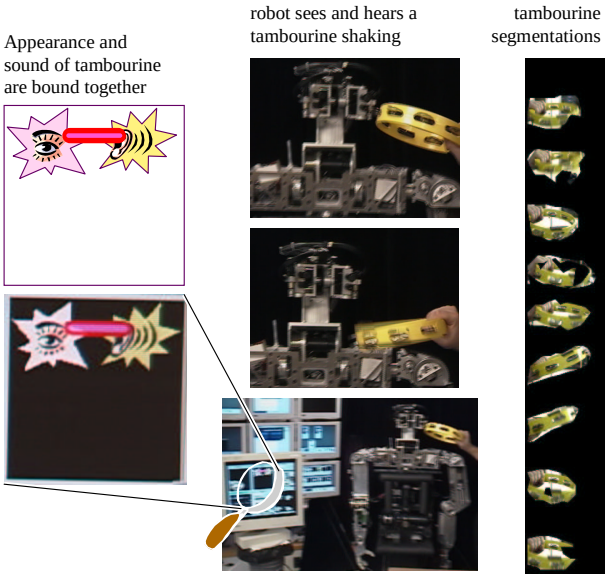

Figure 5: Here the robot is shown a tambourine in use. The robot detects that there is a periodically moving visual source, and a periodic sound source, and that the two sources are causally related and should be bound. All images in these figures are taken directly from recordings of real-time interactions, except for the summary box in the top-left (included since in some cases the recordings are of poor quality). The images on the far right show the visual segmentations recorded for the tambourine in the visual modality. The background behind the tambourine, a light wall with doors and windows, is correctly removed. Acoustic segmentations are generated but not shown (see Figures 3 and 4 for examples).

in (Arsenio and Fitzpatrick, 2003) and is illustrated in Figure 4). For these examples and all other experiments described in this paper, our system tracks moving pixels in a sequence of images from one of the robot's cameras using a multiple tracking algorithm based on a pyramidal implementation of the Lukas-Kanade algorithm. A microphone array samples the sounds around the robot at 16kHz. The Fourier transform of this signal is taken with a window size of 512 samples and a repetition rate of 31.25Hz. The Fourier coefficients are grouped into a set of frequency bands for the purpose of further analysis, along with the overall energy.

4. Learning about objects

Segmented features extracted from visual and acoustic segmentations (using the method presented in last section) can serve as the basis for an object recognition system. In the visual domain, (Fitzpatrick, 2003) used segmentations derived through physical contact as an opportunity for a robot to become familiar with the appearance of objects in its environment and grow to recognize them. (Krotkov et al., 1996) has looked at recognition of the sound generated by a single contact event. Visual and acoustic cues are both individually important for recognizing objects, and can complement each other when, for example, the robot hears an object that is outside its view, or it sees an object at rest. But when both visual and acoustic cues are present, then we can do even better by looking at the relationship between the visual motion of an object and the sound it generates. Is there a loud bang at an extreme of the physical trajectory? If so we might be looking at a hammer. Are the bangs at either extreme of the trajectory? Perhaps it is a bell. Such relational features can only be defined and factored into recognition if we can relate or *bind* visual and acoustic signals.

Several theoretical arguments support the idea of binding by temporal oscillatory signal correlations (von der Malsburg, 1995). From a practical perspective, repetitive synchronized events are ideal for learning since they provide large quantities of redundant data across multiple sensor modalities. In addition, as already mentioned, extra information is available in periodic or locally-periodic signals such as the period of the signal, and the phase relationship between signals from different senses – so for recognition purposes the whole is greater than the sum of its parts.

Therefore, a binding algorithm was developed to associate cross-modal, locally periodic signals, by which we mean signals that have locally consistent periodicity, but may experience global drift and variation in that rhythm over time. In our system, the detection of periodic cross-modal signals over an interval of seconds using the method described in the previous section is a necessary (but not sufficient) condition for a binding between these signals to take place. We now describe the extra constraints that must be met for binding to occur.

For concreteness assume that we are comparing a visual and acoustic signal. Signals are compared by matching the cluster centers determined as in the previous section. Each peak within a cluster from the visual signal is associated to a temporally close (within a maximum distance of half a visual period) peak from the acoustic signal, so that the sound peak has a positive phase lag relative to the visual peak. Binding occurs if the visual period matches the acoustic one, or if it matches half the acoustic period, within a tolerance of $60ms$. The reason for the second match is that often sound is generated at the fastest points of an object's trajectory, or the extremes of a trajectory, both of which occur twice for every single period of the trajectory. Typically there will be several redundant matches that lead to binding within a window of the sensor data for which several sound/visual peaks were detected. In

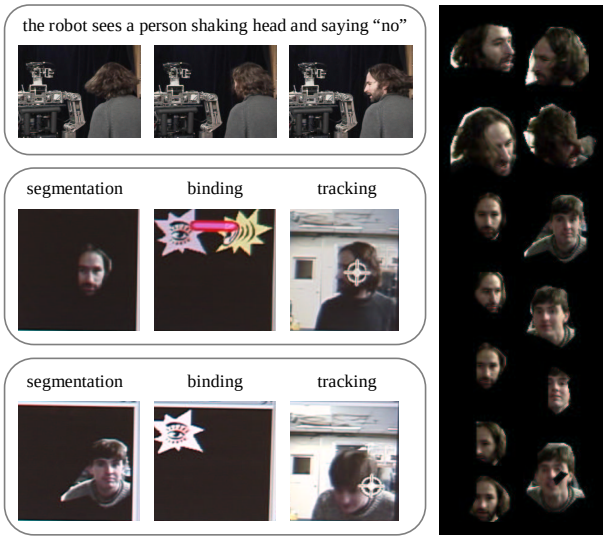

Figure 6: In this experiment, the robot sees people shaking their head. In the top row, the person says "no, no, no" in time with his head-shake. The middle row shows the recorded state of the robot during this event – it binds the visually tracked face with the sound spoken. The lower row shows the state during a control experiment, when a person is just nodding and not saying anything. Recorded segmentations for these experiments are shown on the right.

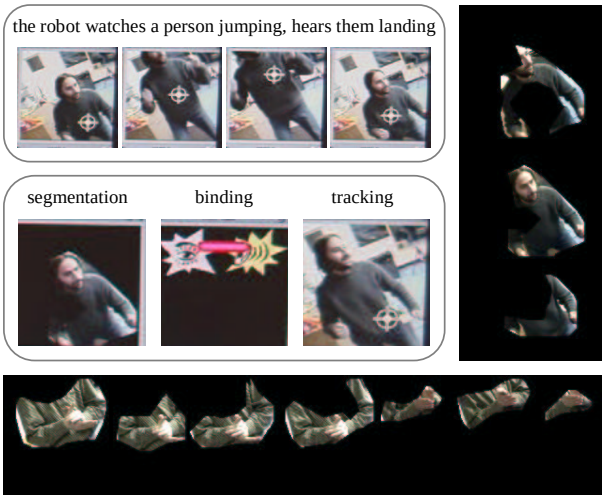

Figure 7: Once the cross-modal binding system was in place, the authors started to have fun. This figure shows the result of one author jumping up and down like crazy in front of the robot. The thud as he hit the floor was correctly bound with segmentations of his body (column on right). The bottom row shows segmentations from a similarly successful experiment where the other author started applauding the robot.

(Arsenio and Fitzpatrick, 2003), we describe a more sophisticated binding method that can differentiate causally unconnected signals with periods that are similar just by coincidence, by looking for a drift in the phase between the acoustic and visual signal over time, but such nuances are less important in a benign developmental scenario supported by a caregiver.

Figure 5 shows an experiment in which a person shook a tambourine in front of the robot for a while. The robot detected the periodic motion of the tambourine, the rhythmic rise and fall of the jangling bells, and bound the two signals together in real-time.

5. Learning about people

In this section we do not wish to present any new algorithms, but rather show that the cross-modal binding method we developed for object perception also applies to perceiving people. Humans often use body motion and repetition to reinforce their actions and speech, especially with young infants. If we do the same in our interactions with Cog, then it can use those cues to link visual input with corresponding sounds. For example, Figure 6 shows a person shaking their head while saying "no! no! no!" in time to his head motion. The figure shows that the robot extracts a good segmentation of the shaking head, and links it with the sound signal. Such actions appear to be understood by human infants at around 10-12 months (American Academy Of Pediatrics, 1998).

Sometimes a person's motion causes sound, just as an ordinary object's motion might. Figure 7 shows a person jumping up and down in front of Cog. Every time he land on the floor, there is a loud bang, whose periodicity matches that of the tracked visual motion. We expect that there are many situations like this that the robot can extract information from, despite the fact that those situations were not considered during the design of the binding algorithms. The images in all these figures are taken from online experiments – no offline processing is done.

6. Learning about the self

So far we have considered only external events that do not involve the robot. In this section we turn to the robot's perception of its own body. Cog treats proprioceptive feedback from its joints as just another sensory modality in which periodic events may occur. These events can be bound to the visual appearance of its moving body part – assuming it is visible – and the sound that the part makes, if any (in fact Cog's arms are quite noisy, making an audible "whirr-whirr" when they move back and forth).

Figure 8 shows a basic binding experiment, in which a person moved Cog's arm while it is out of the robot's view. The sound of the arm and the robot's proprioceptive sense of the arm moving are bound together. This is an important step, since in

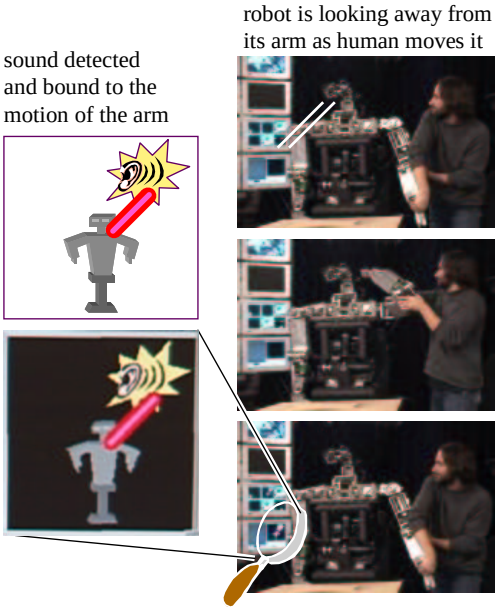

Figure 8: In this experiment, a person grabs Cog's arm and shakes it back and forth while the robot is looking away. The sound of the arm is detected, and found to be causally related to the proprioceptive feedback from the moving joints, and so the robot's internal sense of its arm moving is bound to the external sound of that motion.

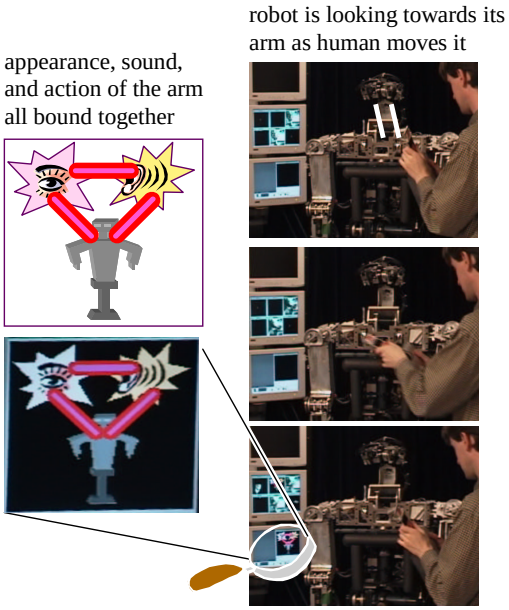

Figure 9: In this experiment, a person shakes Cog's arm in front of its face. What the robot hears and sees has the same rhythm as its own motion, so the robot's internal sense of its arm moving is bound to the sound of that motion and the appearance of the arm.

the busy lab Cog inhabits, people walk into view all the time, and there are frequent loud noises from the neighboring machine shop. So cross-modal rhythm is an important cue for filtering out extraneous noise and events of lesser interest.

In Figure 9, the situation is similar, with a person moving the robot's arm, but the robot is now looking at the arm. In this case we see our first example of a binding that spans three modalities: sight, hearing, and proprioception. The same is true in Figure 10, where Cog shakes its own arm while watching it in a mirror. This idea is related to work in (Metta and Fitzpatrick, 2003), where Cog located its arm by shaking it.

An important milestone in child development is reached when the child recognizes itself as an individual, and identifies its mirror image as belonging to itself (Rochat and Striano, 2002). Self-recognition in a mirror is also the focus of extensive study in biology. Work on self-recognition in mirrors for chimpanzees (Gallup et al., 2002) suggests that animals other than humans can also achieve such competency, although the interpretation of such results requires care and remains controversial. Self-recognition is related to the notion of a theory-of-mind, where intents are assigned to other actors, perhaps by mapping them onto oneself, a topic of great interest in robotics (Kozima and Yano, 2001, Scassellati, 2001). Proprioceptive feedback provides very useful reference signals to identify appearances of the robot's body in different modalities. That is why we extended our binding algorithm to include proprioceptive data.

Children between 12 and 18 months of age become interested in and attracted to their reflection (American Academy Of Pediatrics, 1998). Such behavior requires the integration of visual cues from the mirror with proprioceptive cues from the child's body. As shown in Figure 11, the binding algorithm was used not only to identify the robot's own acoustic rhythms, but also to identify visually the robot's mirror image (an important milestone in the development of a child's theory of mind (Baron-Cohen, 1995)). It is important to stress that we are dealing with the low-level *perceptual* challenges of a theory of mind approach, rather than the high-level *inferences* and mappings involved. Correlations of the kind we are making available could form a grounding for a theory of mind and body-mapping, but are not of themselves part of a theory of mind – for example, they are completely unrelated to the intent of the robot or the people around it, and intent is key to understanding others in terms of the self (Kozima and Zlatev, 2000, Kozima and Yano, 2001). Our hope is that the perceptual and cognitive research will ultimately merge and give a truly intentional robot that understands others in terms of its own goals and body image – an image which could develop incrementally using cross-modal correlations of the kind explored in this paper.

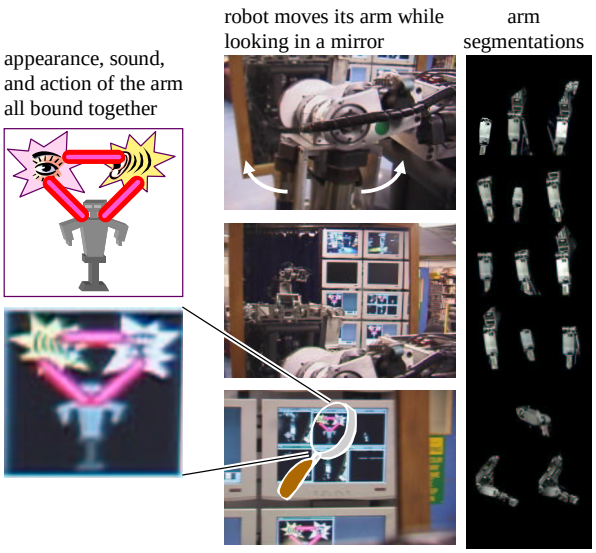

Figure 10: In this experiment, Cog is looking at itself in a mirror, while shaking its arm back and forth (the views on the right are taken by a camera behind the robot's left shoulder, looking out with the robot towards the mirror). The reflected image of its arm is bound to the robot's sense of its own motion, and the sound of the motion. This binding is identical in kind to the binding that occurs if the robot sees and hears its own arm moving directly without a mirror. However, the appearance of the arm is from a quite different perspective than Cog's own view of its arm.

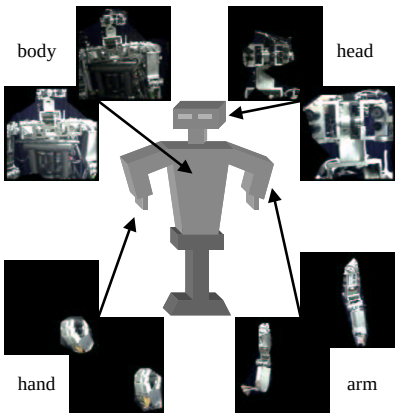

Figure 11: Cog can be shown different parts of its body simply by letting it see that part (in a mirror if necessary) and then shaking it, such as its (right) hand or (left) flipper. Notice that this works for the head, even though shaking the head also affects the cameras.

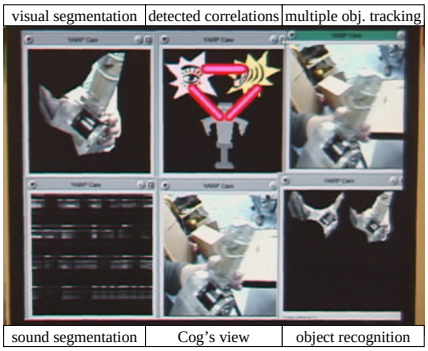

Figure 12: This figure shows a real-time view of the robot's status during the experiment in Figure 9. The robot is continually collecting visual and auditory segmentations, and checking for cross-model events. It also compares the current view with its database and performs object recognition to correlate with past experience (bottom right).

7. Discussion and conclusions

Most of us have had the experience of feeling a tool become an extension of ourselves as we use it (see (Stoytchev, 2003) for a literature review). Many of us have played with mirror-based games that distort or invert our view of our own arm, and found that we stop thinking of our own arm and quickly adopt the new distorted arm as our own. About the only form of distortion that can break this sense of ownership is a delay between our movement and the proxy-arm's movement. Such experiences argue for a sense of self that is very robust to every kind of transformation except latencies. Our work is an effort to build a perceptual system which, from the ground up, focuses on timing just as much as content. This is powerful because timing is truly cross-modal, and leaves its mark on all the robot's senses, no matter how they are processed and transformed.

We are motivated by evidence from human perception that strongly suggests that timing information can transfer between the senses in profound ways. For example, experiments show that if a short fragment of white noise is recorded and played repeatedly, a listener will be able to hear its periodicity. But as the fragment is made longer, at some point this ability is lost. But the repetition can be heard for far longer fragments if a light is flashed in synchrony with it (Bashford et al., 1993) – flashing the light actually changes how the noise sounds. More generally, there is evidence that the cues used to detect periodicity can be quite subtle and adaptive (Kaernbach, 1993), suggesting there is a lot of potential for progress in replicating this ability beyond the ideas already described.

Although there is much to do, from a practical perspective a lot has already been accomplished. Consider Figure 12, which shows a partial snapshot of the robot's state during one of the experiments described in the paper. The robot's experience of an event is rich, with many visual and acoustic segmentations generated as the event continues, relevant prior segmentations recalled using object recognition, and the relationship between data from different senses de-

tected and stored. We believe that this kind of experience will form one important part of a perceptual toolbox for autonomous development, where many very good ideas have been hampered by the difficulty of robust perception.

Another ongoing line of research we are pursuing is truly cross-modal object recognition. A hammer causes sound after striking an object. A toy truck causes sound while moving rapidly with wheels spinning; it is quiet when changing direction – therefore, the car's acoustic frequency is twice as much as the frequency of its visual trajectory. A bell typically causes sound at either extreme of motion. All these statements are truly cross-modal in nature, and with our system we can begin to use such properties for recognition.

Acknowledgements

Project funded by DARPA as part of the "Natural Tasking of Robots Based on Human Interaction Cues" under contract number DABT 63-00-C-10102, and by the Nippon Telegraph and Telephone Corporation as part of the NTT/MIT Collaboration Agreement. Author supported by Portuguese grant PRAXIS XXI BD/15851/98.

References

American Academy Of Pediatrics (1998). *Caring for Your Baby and Young Child: Birth to Age 5*. Bantham.

Arsenio, A. and Fitzpatrick, P. (2003). Exploiting cross-modal rhythm for robot perception of objects. In *Proceedings of the Second International Conference on Computational Intelligence, Robotics, and Autonomous Systems*, Singapore.

Arsenio, A., Fitzpatrick, P., Kemp, C. C., and Metta, G. (2003). The whole world in your hand: Active and interactive segmentation. Proceedings of the Third International Workshop on Epigenetic Robotics.

Baron-Cohen, S. (1995). *Mindblindness*. MIT Press.

Bashford, J. A., Brubaker, B. S., and Warren, R. M. (1993). Cross-modal enhancement of repetition detection for very long period recycling frozen noise. *Journal of the Acoustical Soc. of Am.*, 93(4):2315.

Brooks, R. A., Breazeal, C., Marjanovic, M., Scassellati, B., and Williamson, M. M. (1999). The Cog project: Building a humanoid robot. In Nehaniv, C. L., (Ed.), *Computation for Metaphors, Analogy and Agents*, volume 1562 of *Springer Lecture Notes in Artificial Intelligence*, pages 52–87. Springer-Verlag.

Fitzpatrick, P. (2003). Object lesson: discovering and learning to recognize objects. Proceedings of the Third International Conference on Humanoid Robots, Karlsruhe, Germany.

Gallup, G., Anderson, J. R., and Shillito, D. J. (2002). The mirror test. In Bekoff, M., Allen, C., and Burghardt, G. M., (Eds.), *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, pages 325–33. Bradford Books.

Hendriks-Jansen, H. (1996). *Catching Ourselves in the Act*. MIT Press, Cambridge, Massachusetts.

Kaernbach, C. (1993). Temporal and spectral basis of the features perceived in repeated noise. *Journal of the Acoustical Soc. of Am.*, 94(1):91–97.

Kozima, H. and Yano, H. (2001). A robot that learns to communicate with human caregivers. In *Proceedings of the First International Workshop on Epigenetic Robotics*.

Kozima, H. and Zlatev, J. (2000). An epigenetic approach to human-robot communication. In *IEEE International Workshop on Robot and Human Communication (ROMAN00)*, Osaka, Japan.

Krotkov, E., Klatzky, R., and Zumel, N. (1996). Robotic perception of material: Experiments with shape-invariant acoustic measures of material type. In *Experimental Robotics IV*. Springer-Verlag.

Metta, G. and Fitzpatrick, P. (2003). Early integration of vision and manipulation. *Adaptive Behavior*, 11(2):109–128.

Rochat, P. and Striano, T. (2002). Who's in the mirror? self-other discrimination in specular images by four- and nine-month-old infants. *Child Development*, 73(1):35–46.

Scassellati, B. (2001). *Foundations for a Theory of Mind for a Humanoid Robot*. PhD thesis, MIT Department of Electrical Engineering and Computer Science.

Stoytchev, A. (2003). Computational model for an extendable robot body schema. Technical report, Georgia Institute of Technology, College of Computing. GIT-CC-03-44.

von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology*, 5:520–526.