

Event Perception: Motor Theory vs. Direct Perception Theory

Young-Lim Lee¹

Abstract

People observe lots of events around the environment and we can easily recognize the nature of an event from the resulting optic flow. The questions are how do people recognize events and what is the information in the optic flow that enables observers to recognize events. There have been debates between motor theorists and direct perception theorists regarding of these problems. Motor theorists claim that human observers exhibit special sensitivity when perceiving speech or biological motion, because we both produce and perceive those events. However, direct perception theorists suggested that speech or biological motion is not special from the perception of all other kinds of event. In this article, I review this controversy to critique the motor theory and to describe a direct realist approach to event perception.

In everyday life, we observe lots of events around us, such as a person walking, a ball bouncing, water falling, etc. There is no problem for us to recognize the nature of an event from the resulting optic flow. Researchers have been investigating the questions of how people recognize events and what information is in the optic flow that enables observers to recognize events. Bingham (1995) suggested that events could be characterized in terms of dynamics. The perceptual information is projected from kinematics or motions in an event. Kinematic specification of dynamics (KSD) was formulated as a principle to be used to guide investigations of perceptual information and suggested that dynamics enables a kinematic pattern to specify events (Runeson, 1977; Runeson & Frykholm, 1983).

¹ Dankook University, Requests for reprints should be sent to Young-Lim Lee, Psychology department, Dankook University, 119 Dandae-ro, Dongnam-gu, Cheonan-si, Chungnam, South Korea.

* This work was supported by a grant from Dankook University (No. 122366).

For example, in terms of dynamics, walking can be understood as an upright and inverted pendulum, while running can be understood as a bouncing ball (Alexander, 1992).

There have been debates between motor theorists and direct perception theorists. According to motor theorists, speech and biological motion are different from other events because humans both produce and perceive those events. As a result, motor theorists claim that human observers exhibit special sensitivity when perceiving speech or biological motion. However, Fowler (1986, 1989) critiqued the motor theory of speech perception and suggested that speech perception is not special. Instead, she advocated a direct realist approach to the perception of speech. In recent papers, researchers in visual event perception have argued for a motor theory of biological motor perception. The problem with this approach is that it divorces biological event perception from the perception of all other kinds of events. In this article, I review this controversy to critique the motor theory and to describe a direct realist approach to event perception.

I. What is Event Perception? Johansson vs. Gibson

What makes events distinguished from other physical properties, such as 3D structure, is that they are spatio-temporal (Warren & Shaw, 1985). One of the prominent examples of event perception is point-light walkers in Johansson's study (1973). Small lights were attached to the main joints of an actor dressed in dark clothes, thus only several small lights were seen in the displays. In the static displays, only a jumble was seen, but in the changing displays, a motion pattern such as walking or running was recognized. Whenever the changing displays were stopped, a jumble was seen again.

Besides Johansson, Gibson is another dominant event perception theorist. Although Johansson is a motor theorist, Gibson is a direct perception theorist. Mace (1985) pointed out the salient differences between the two positions. First, Johansson and Gibson had different viewpoints on events whose changes are slow or fast. Johansson (1970) distinguished perceptible events from nonperceptible events. He said that humans do not perceive slow motion such as ripening of fruits or the fast motion of a raster which generates a static TV image. Those events change so slow or fast that they are not sufficient to be perceived and recognized.

On the other hand, Gibson (1970, 1979) proposed that the speed of an event is not important, rather the existence and availability of information is important to perceive events. He described the perception as the pickup of information over time, thus both slow and fast events might be perceived as long as information specifying them is available to the observer.

Second, Johansson and Gibson devised different displays that produced the changing patterns over static ones. As I mentioned briefly before, Johansson (1973, 1985) developed the vector analysis using the point-light method. For instance, a person slowly passes a factory wall when a piece of wallboard is being lifted in front of the wall. The wallboard motion can be described as the vector sum of the projection of a horizontal motion from the wall relative to the eye and a vertical motion from the board relative to the wall. In other words, in the optic flow projected at the retina, the horizontal vectors from the wall indicate a horizontal relative motion between the eye and the wall and the diagonal vectors from the rectangle wallboard indicate a diagonal relative motion between the eye and the wallboard. Thus, the perception of two simultaneous motions is the results of mathematical analysis in sensory processing. (Johansson, 1985) Gibson (1979), however, devised random texture displays to study the specification and transformation of surfaces. He progressively added or subtracted textures from each successive frame that looked like opaque surfaces moving over one another. He emphasized the information available about a cluttered layout of surfaces in the addition and subtraction of textures. Since the changes from accretion and deletion of texture are not found in projective geometry, he argued that projective geometry could not be the most general source for the specification and transformation of surfaces.

Third, the most significant difference between Johansson and Gibson is the underlying paradigms of perception. Johansson (1985) adapted the distal stimulus–proximal stimulus–percept paradigm. A proximal stimulus is formed by light (i.e., distal stimulus) reflected from the environment and specified at the optic nodal point of the eye. This specified proximal distribution of energy is transformed into a percept by neural processing in the visual system. Gibson (1972), on the other hand, questioned the idea that perception is mediated by stimulus in distal-proximal framework and proposed that perception is directly produced by structure (i.e., pattern) in an environment. From his view, the information is the pattern (optic, acoustic, haptic, etc.) specific to its sources.

In the case of vision perception, light is structured differently by different objects and events in the environment. In other words, structure in light is caused by properties of objects and events and the structure is specific to its causal source. The structure in light which serves as information is directly picked up to perceive the surface layout of the environment.

So far, only the differences between two major event perception theorists have been briefly reviewed in terms of the basic concept of event perception. In the following sections, I review the controversy between motor theory and direct perception theory more specifically in the areas of speech and biological motion perception.

II. Speech Perception

1. Introduction of Speech Perception

From both motor theory and direct perception theory perspectives, the objects of speech perception are articulatory, rather than auditory or acoustic (Diehl & Kluender, 1989). Individual phonetic segments (i.e., vowels and consonants) are realized articulatorily as sets of coordinated gestures of various vocal structures such as tongue body and tip, lips, soft palate, jaw, and larynx (Fowler, 1986; also see Kelso, Tuller, Vatikiotis-Bateson & Fowler, 1984). Speech perception is dynamic as well as other events. Remez, Rubin, Pisoni, and Carrell (1981) showed that time-varying properties of artificial acoustic signals are sufficient to provide information to perceive a linguistic message in speech, even though their stimuli had no acoustic elements for phonetic segments. Their finding is similar as Johansson's (1973) point-light demonstrations in which several lights are recognized as a motion pattern when changed over time (Folwer & Rakerd, 1985). Although motor theorists and direct perception theorists agree with the notion that the objects of speech perception are articulatory (i.e., dynamic), rather than auditory or acoustic, they have been in debate to solve the question whether aspects of speech perception are or are not special compared to perception of nonspeech sound.

2. Motor Theory

Motor theory proposes that speech perception is special. Humans are more sensitive to speech relative to nonspeech sounds because only speech sounds can be perceived and produced. The perceived speech sounds are continuously articulated and compared with the auditory result of the articulation (Denes & Pinson, 1996).

Liberman and Mattingly (1985) have presented a comprehensive theory, in which articulatory commands play a major role in speech perception.

First, "the objects of speech perception are the intended phonetic gestures of the speaker, represented in the brain as invariant motor command that call for movements of the articulators through certain linguistically significant configurations (p. 2)." Phonetic segments consist of one or more articulatory gestures such as tongue backing, lip rounding, and jaw raising. To perceive speech sounds is to perceive a specific pattern of intended gestures. Second, "[since] speech perception and speech production share the same set of invariants, they must be intimately linked. The link is not a learned association, rather [it] is innately specified. (p.3)" Perception of sounds is specialized by phonetic gestures depending on articulator movements and vocal-tract shapes. Thus, speech perception is special because the link between perception and production occurs only in speech. As Liberman and Mattingly claimed, the link between perception and production is innately specified. Eimas, Siqueland, Jusczyk, and Vigorito (1971) found that 1- and 4-month-old infants distinguished the acoustic cues between the voiced and voiceless stop consonants /b/ and /p/ much as adults did. The difference between /b/ and /p/ is voice onset time (VOT), defined as the time between the release burst and the onset of voicing; /b/ has a short voiced lag, whereas /p/ has a long voiced lag. Infants were habituated to the first speech sound, and then their response rate to a second speech sound was measured to investigate their ability to discriminate acoustic sounds within the phonemic categories on the basis of VOT. Infants showed greater recovery from habituation when the two stimuli were from different phonemic categories than when they were from same phonemic categories. It has been shown that prelinguistic infants were able to discriminate voiced stops from voiceless stop consonants on the basis of VOT because perception and production link is not associated with learning, but specified innately.

The fundamental notion of motor theory is that humans have a special sensitivity to perceive speech sounds relative to nonspeech sounds. Researchers have investigated whether speech perception is distinguished from perception of nonspeech analog (Diehl & Walsh, 1989; Liberman, Harris, Kinney & Lane, 1961; Pisoni, Carrell & Gans, 1983).

Lieberman et al. (1961) investigated the discriminability of speech sounds compared to nonspeech sounds. They generated the spectrograms used to produce the stimuli /do/ and /to/ using the pattern playback which converts the spectrograms into sound. The distinction of speech stimuli between /do/ and /to/ is that these patterns differ only in the relative time of onset of the first formant relative to the second and third formants. Nonspeech stimuli were simply made by turning the speech spectrograms upside down. Although nonspeech stimuli are not perceived as speech, essentially the same acoustic differences (i.e., the relative time of onset of the formants) occur in nonspeech stimuli as speech stimuli. Thus, nonspeech stimuli can be discriminated if observers use only the relative onset time. Observers heard three stimuli and asked to decide whether the third stimulus was identical to the first or the second one. The results showed that observers discriminated nonspeech sounds much more poorly than speech sounds. Thus, Liberman et al. concluded that speech and nonspeech sounds are dissimilar.

Although Liberman et al. (1961) found that perceived speech sounds are different from nonspeech sounds, somewhat different results have been found in other experiments. Pisoni et al. (1983) and Diehl and Walsh (1989) investigated the distinction between stops and glides (e.g., /b/ vs. /w/) by comparing speech stimuli, /ba/ and /wa/, and corresponding nonspeech stimuli. According to the auditory principle of durational contrast, the duration of acoustic segments affects the perception of length of an adjacent segment (Miller & Liberman, 1979). For instance, a longer vowel will produce shorter formant transitions, thus, more stop identification responses. In both studies, they found that frequency transition duration is an effective cue to distinguish speech sounds, stops and glides, as well as nonspeech sounds, abrupt and gradual onsets. In other words, there was no difference between speech and nonspeech distinction on the basis of transition duration. However, Diehl and Walsh found that speech and nonspeech sounds are dissimilar with respect to amplitude rise time, although they are similar with respect to transition duration. When amplitude rise time was varied while transition duration was fixed, variation in rise time had only a small effect on distinction between speech (stops vs. glides) as well as nonspeech (abrupt vs. gradual onset) sounds. However, speech sounds showed a stimulus length effect, in which a longer vowel shifted the boundary of stop/glide distinction toward being longer, hence, glide sounds were identified more as stop sounds.

In comparisons of speech with nonspeech sounds, findings are not consistent with the motor theory's claim that speech perception is special. Motor theorists explain the divergent findings in two ways. One possibility is that the ability to perceive differences among the sounds of speech is constrained by the processes, that is, some processes applied to speech sounds are special while others are not (e.g., Eimas, 1985). Another possibility is that some nonspeech sounds are so speechlike as to be perceived as speech while others are not (e.g., Diehl & Walsh, 1989; Best, Studdert-Kennedy, Manuel & Rubin-Spitz, 1989). These possibilities, however, cannot be sufficient to support the claim that speech perception is special. Direct perception theorists, thus, claim that speech perception is not special.

3. Direct Perception Theory

In contrast to motor theory's claim that speech signals are perceived by special inner mechanisms to produce speech, direct perception theory denies the special perception-production link (Fowler & Galantucci, 2005). Since the same acoustic signals could be formed by different vocal-tract configurations and the same vocal-tract configurations could be used for different acoustic signals, motor theorists claim that speech signals are internally computed by the nervous system to be perceived (Liberman & Mattingly, 1985). In other words, the speech motor system is used in perception to help extract articulatory movements which produce the acoustic speech signal together. On the other hand, as Gibson (1972) proposed, direct perception theorists claim that structure in sine waves is directly picked up to perceive the acoustic signals. There is no distinction between speech and nonspeech events. Motor theorists infer that different perceptual processes are applied to the acoustic signals, one to speech and one to nonspeech because observers respond differently to speech and nonspeech signals (e.g., Liberman et al., 1961). Direct perception theorists, on the other hand, suggest that responses to speech and nonspeech signals do not occur due to the different perceptual processes. Instead, responses to acoustic signals are affected by what the signals are perceived as (Fowler, 1990). In other words, acoustic signals are perceived depending on directly picked up information in the environment no matter what signals are, speech or nonspeech.

Fowler (1990) investigated direct perception theory's claim that speech perception is not special by using nonspeech signals similar to the ba-wa stimuli of Miller and Liberman (1979).

As I mentioned earlier, Pisoni et al. (1983) and Diehl and Walsh (1989) investigated whether observers respond to stop/glide speech and abrupt/gradual onset nonspeech stimuli differently. Pisoni et al. found that responses to speech and nonspeech stimuli are similar, and thus that perceivers use durational information in speech as well as nonspeech signals. Diehl and Walsh replicated the Pisoni et al.'s comparison study between speech and nonspeech analogs because they were concerned that the nonspeech analogs that Pisoni et al. used might be processed as speech due to three time-varying sine wave segments. Although Diehl and Walsh used a single sine wave segment for nonspeech analogs, they found that speech and nonspeech sounds are similar in respect to transition duration. However, they also found that speech and nonspeech sounds are dissimilar with respect to amplitude rise time. To ascribe the similarity between speech and nonspeech signals, Fowler used nonspeech events similar to the ba/wa stimuli instead of the synthesized nonspeech analogs that Pisoni et al. and Diehl and Walsh used. Nonspeech events she used were produced by a steel ball rolling down a set of steel tracks. Sounds were recorded as a ball rolling down from the downward slopes onto the flat or upward sloping tracks.

Each event consisted of the phase 1 sound (i.e., durations of the downward slopes) followed by the phase 2 sound (i.e., durations of either the flat or upward sloping track). Two phases of the event were stored separately and constructed by splicing different phase 2 sounds onto the five phase 1 sounds. The phase 1 sound was produced by one of five downward slopes at 50, 40, 30, 20, and 10 degrees relative to the horizontal and the phase 2 sound was produced by either 10 or 50 degree tracks. A steeper slope is associated with shorter duration in both phases. Durations in phase 2 are positively or negatively related to those in phase 1. In the event with the flat track, duration in phase 2 is longer as the slope of phase 1 is shallower (i.e., 10 degrees). That is, durations in phase 2 are positively related to those in phase 1, since a long-duration phase 2 implies a long-duration phase 1. In the event with the upward sloping track, on the other hand, duration in phase 2 is longer as the slope of phase 1 is steeper (i.e., 50 degrees). That is, durations in phase 2 are negatively related to those in phase 1, since a long-duration phase 2 implies a short-duration phase 1. The downward sloping part (i.e., phase 1) of each track was covered with sandpaper so that the sound in phase 2 changed noticeably as the steel ball rolled onto the flat (or upward slope) from the sloping part of the track. Observers were asked to judge the slope of the downward ramp (phase 1 part) of each track by choosing between the classifications steep or shallow.

Fowler predicted that if acoustic signals were used as information to be perceived directly, rather than through perceptual processes, judgments of phase 1 slopes would be affected by the durations of phase 2. The results were consistent with the prediction.

In the upsloping condition, observers judged the slopes of phase 1 steeper (i.e., short duration) when followed by the long (50 degree) duration phase 2 than by the short (10 degree) duration phase 2. In the flat condition, on the other hand, observers judged the slopes of phase 1 steeper when followed by the short (50 degree) duration phase 2 than by the long (10 degree) duration phase 2. When the duration of phase 2 was short, there was difference in perception of phase 1 between two conditions. Since the relation between phase 1 and 2 is positive in the flat condition, durations of phase 1 were perceived shorter in this condition. Since the relation between phase 1 and 2 is negative in the upsloping condition, on the other hand, durations of phase 1 were perceived longer in this condition. The findings of this study, that the durations of phase 2 were used as information for the slope of a ramp causing the phase 1 sound structure, are similar to the durational contrast effect that Miller and Liberman found in speech perception. Thus, there is no difference between speech and nonspeech sounds because listeners use structure in the acoustic signals as information for its causal source in perceiving nonspeech sounds as well as speech sounds.

III. Biological Motion Perception

Speech perceptionists have been in a debate. Basically, motor theorists claim that speech perception is special because the perceptual-motor mechanism (i.e., production of speech) allows humans have great sensibility to perceive speech relative to nonspeech signals. Direct perception theorists, on the other hand, claim that there is no distinction between speech and nonspeech signals specified by different perceptual processes. Instead, humans directly use the structure in the acoustic signals as the information to perceive speech as well as nonspeech. More recently, biological motion perceptionists have repeated this debate on motor theory. Similarly to the debate of speech perceptionists, motor theorists in biological motion perception have argued that human movements are special, whereas direct perception theorists have critiqued this motor theory's claim.

1. Motor Theory

From Johansson's point-light demonstration, motor theorists claim that body movements can be recognized easily even when they are depicted visually in a simple form by placing a few lights at the articulations of the body segments (Viviani & Stucchi, 1992). For example, biological motions, such as walking (Beardworth & Bukner, 1981; Cutting, 1981; Jacobs & Pinto, 2004; Jacobs & Shiffrar, 2005; Johansson, 1973; Loula, Prasad, Harber & Shiffrar, 2005), dancing (Johansson, 1973; 1977), and other motions (Loula et al., 2005) have been investigated to support this claim. Beardworth and Bukner (1981) performed an experiment on the ability to recognize walking movements. They adapted the point-light displays of Johansson and recorded the dynamic walking of each observer and his/her friends whose walking was familiar to the observer because they spent time every day. After being shown each walking display of their own as well as their friends', observers were asked to indicate the name of the walker. The results showed that performances on recognition of their own walking were better than those of their friends' walking, even though the observer sees the walking of their friends every day, whereas they barely see their own walking. Based on this result, researchers suggested that some specific motor information rather than visual experiences is involved in perceptual recognition.

Jacobs and Pinto (2004) examined the roles of visual experience and motor processes in biological motion perception. They used point-light displays by varying motor feasibility (possible vs. impossible) and gait type (familiar, unusually fast, unusually slow) and performed gait-speed and identity discrimination tasks. In gait-speed tasks, observers were shown two walking displays and asked to judge which one walked fast and which one reached the door first. In the identity discrimination task, observers were asked to judge whether the displayed walker was model 1 or 2. In both tasks, the accuracy of judgments for physically possible gaits was higher than for physically impossible gaits, meaning that perceptual-motor interaction plays a role in human movement perception. Their results are consistent with the neural reports of Stevens, Fonlupt, Shiffrar, and Decety (2000). Stevens et al. investigated brain activity of the motor and parietal cortex using PET when biomechanically possible and impossible stimuli were presented. They found that the motor and parietal region was no longer activated when impossible paths of human movement were presented. Jacobs and Pinto also tested the role of visual experience. Observers were shown two walker models for a varied time (i.e., more than 20 hours vs. less than 5 hours).

Visual experience affected the identity recognition, but not gait-speed discrimination. Thus, visual experience as well as perceptual-motor interaction allows us have a great sensitivity to human movements.

Since perceptual-motor interaction defines visual sensitivity to human movements, Jacobs and Shiffrar (2005) investigated whether different motions of observers (e.g., stationary, walking or cycling observers) affects different judgments of gait speed on self- and other-relative movements. Observers were asked to judge either whether their own walking speed was faster than a point-light walker's or which one walked faster between two point-light walkers while they stood, walked, or bicycled. They found that the accuracy of discrimination of walking speed was lower when observers walked than when observers stood or bicycled. Researchers suggested that action production interferes with action perception. Observer's own walking interfered with the sensitivity to perceive walking movements.

Motor theorists claim that human movements are special, meaning that humans have a great sensitivity to human movements. Moreover, they argue that the perceptual-motor mechanism plays a role in perception of biological motions. Since humans both produce and perceive motions, observers recognize their own motions better than others' even though they have little visual experience of their own motions relative to others' motions. Thus, Loula et al. (2005) also predicted that observers should recognize their own motion better than the motion of friends or strangers in their experiments. They investigated several kinds of human movements such as walking, running, dancing, boxing, jumping, hugging, greeting, and laughing. Their results were somewhat consistent with their prediction because observers recognized themselves better than friends and strangers for certain kinds of motions. For example, observers recognized their own dancing better than others', but they did not recognize their own walking and running better than others'. Thus, when the movements are biomechanically stereotypical (e.g., walking and running), observers could not discriminate their own movements and others' easily.

2. Critique of the Motor Theory for Biological Motion Perception

Although point-light demonstrations shown in numerous studies are powerful and effective and have advanced motor theory's claim that biological motion perception is special, direct perceptionists have critiqued motor theory's claim of dissociation of biological motion and other object motion perception (c.f., Bingham & Wickelgren, 2008). First, direct perception theorists claim that there is no clear boundary between biological motion and other object motion. A good example of this claim is a running basketball player who is dribbling a ball for a successful lay-up.

Alexander (1992) defined human biological motions in terms of dynamics. Walking can be understood as an upright combined with inverted pendulum, while running can be understood as a bouncing ball. Thus, the dynamic of running motion is same as the dynamic of a bouncing ball. The basketball player produces dribbling, but the dribbling is a part of the bouncing ball. Warren, Kim and Husney (1987) investigated the ability to perceive the elasticity of a ball while observing another person bounce it. They varied elasticities of balls by filling the balls with various materials such as foam rubber, cloth, or styrofoam clips, and by adjusting air pressure. In other respects including weight, size, and color, every ball was identical. Observers were asked to bounce each ball vertically so that the bottom of the ball reached the target line. The experimenter first demonstrated the task and observers dribbled the ball twice, finally bouncing the ball to make it reach the target line. They found that observers were able to perceive the elasticity of the ball by observing how the experimenter bounced it. Moreover, observers were able to use information of the elasticity they perceived to regulate the impulse applied to the ball during bouncing it. Thus, ball bouncing was also an event that humans were able to both perceive and produce. The perception of this event (i.e., ball bouncing) was a single event that combined the biological human limb motions forcing the ball and the inanimate ball bouncing. In a similar vein, the dynamic of the lay-up movement that the player produced exhibits the projectile motion same as the dynamic of thrown ball. Thus, biological motion events (i.e., running and laying-up) are connected to object motion events (i.e., a bouncing ball and projectile motion of a thrown ball). If so, could humans perceive biological motion separately from object motion? Could we say that biological motion is the only category of motion that humans perceive and produce?

Second, motor theorists claim that humans have little visual experience of self motion relative to motions of others. However, people always visually perceive their own motions from a first person perspective.

Bingham and Wickelgren (2008) demonstrated the effective first person display of biological motion. Observers can easily recognize walking movement in a videotaped walking display from a third person perspective. Not only can observers recognize walking motion from the third person perspective, but they can also recognize it in a display videotaped from the first person perspective. For example, when a person holds the video camera on his/her shoulder and films the view while he/she is getting up from a chair and walking around the room, the view through the camera lens is used to guide self motion. The event of walking motion can be easily distinguished from motion of the camera on a tripod, even though there is no such walking motion from the third person perspective.

Self motion is also perceived kinesthetically as well as visually from a first person perspective. Kinesthetic perception is defined as an intrinsic module of the coordination and control of actions (Latash, 1993). Knoblich and Prinz (2001) investigated the role of kinesthetic perception to distinguish self-generated drawing movements from other generated movements. They found that observers were able to recognize their own drawing movements, even though they did not see the process of original drawing. They suggested that the drawing movements were kinesthetically perceived when observers performed them and that this kinesthetic perception reproduced the visual recognition of the self-generated characters. However, they concluded that this kinesthetic perception is encoded by motor system, which is explained by motor theory.

From perceptual theory, the relation between kinesthetic and visual perception focusing on the role of relative phase in an event has also been investigated. "Relative phase is a measure of coordination in human movement. The event perception studies have shown that relative phase is a perceptual variable used for the perception of human movement." (Bingham & Wickelgren, 2008, p. 19) Todd (1983) investigated the role of relative phase in perception of gait. Each stimulus consisted of a pair of three connected line segments which appeared as a pair of human legs. The relative positions and orientations of these segments changed over time in a cycle. The particular configuration of different limb segments was determined by the values of seven variable: the position of the torso (X_t , Y_t), the position of the hip (X_h , Y_h), the angle of the upper leg rotated about the hip, the angle of the lower leg rotated about the knee, and the angle of the foot rotated about the ankle.

The step cycle consisted of three distinct phases called the transfer, the landing, and the drive phase. During these phases, seven variables were varied sinusoidally. Observers were asked to judge whether a display looked more like a walking or running gait. It was found that observers were able to perceive the identity of gaits and the difference of walking and running gaits were primarily determined by the movements of the lower leg. A small bump at the beginning of the step cycle was observed in changes in lower leg angle relative to the horizontal during running, but not during walking. Todd also found that the walking gait can be transformed into running (or vice versa) by adding or subtracting a constant value to the angle of the lower leg over the entire step cycle.

3. Direct Perception Theory

Direct perceptionists claim that biological motion perception is not special because such point-light demonstrations for the other object events (e.g., bouncing balls, free falling, splashing water) are equally effective (Bingham, 1987; Bingham, Rosenblum & Schmidt, 1995; Muchisky & Bingham, 2002). Bingham et al. used patch-light displays for 9 events selected to represent rigid-body dynamics (free fall, pendulum, rolling ball, and struck ball), biodynamics (hand-moved spring and hand-moved pendulum), hydrodynamics (stirred water and splash), and aerodynamics (falling leaves). Observers were asked to describe each event in their own words or circle properties in a list. It was found that observers were able to recognize each event in all tasks. Twardy and Bingham (2002) investigated whether observers are sensitive to violations of energy conservation in free-fall events. They used simulations of balls falling freely and then bouncing on a hard surface varying the elasticity or gravity. Observers were viewed each display and then indicate how each event looked natural. The results showed that observers were sensitive to the effect of elasticity and gravity in free-falling events, although they were more sensitive to the effect of decreasing gravity than to that of increasing gravity.

From direct perception theorists, event perception is yielded by information composed of higher order variables (Gibson, 1979). Direct perception theorists in speech perception claim that structure in sine waves is picked up directly to perceive acoustic signals. In a similar vein, direct perception theorists in biological motion perception claim that structure in surfaces and motions is used as information to specify the event for visual recognition. Motor theorists have been skeptical about the claim of direct perception theorists, that is, higher-order invariants.

They argued that it is impossible to perceive events directly by higher-order invariants. Motor theorists suggested that higher-order variables should be decomposed by the perceptual-motor process. Since perceptual-motor processes for speech / biological motion are different from those for nonspeech / object motion, motor theorists have argued that speech and biological motion perception is special. Direct perception theorists, on the other hand, claim that the dynamic events generate the specified trajectory form to be recognized. In other words, different events are specified by different trajectory forms.

Runeson (1977) suggested that observers detect unique kinematic properties to perceive corresponding dynamic properties of the events, which is called KSD (Kinematic Specification of Dynamics). However, the mapping from dynamic event to kinematic pattern appears to collapse a dimension, which causes a missing dimension problem (Warren & Shaw, 1985). "As the dimension of 'depth' was believed to be lost in a flat retinal image, dynamic variables, such as mass, friction, elasticity, and energy are not present in the kinematic description." (Warren & Shaw, 1985, p. 20) Kinematic variables (e.g., position, velocity, acceleration, etc.) require only the length and time dimensions [L, T], while dynamic variables (e.g., mass, force, stiffness, damping, etc.) require one more dimension, the mass [M]. Since, the mass dimension is missing in kinematics, how is it recovered from kinematics to perceive dynamic events?

Bingham (1995) rejected the missing dimension problem, using the mathematical equation shown below. In the case of a mass-spring oscillator, the dynamic equation is $m(d^2x/dt^2) = -kx$, where m is mass, k is stiffness, x is position, and d^2x/dt^2 is acceleration. This equation includes dynamic variables (m and k) and kinematic variables (x and d^2x/dt^2). "Dimensionally, m and k are [M] and [M/T²], while x and d^2x/dt^2 are [L] and [L/T²], so each term in the equation is dimensionally a force, i.e., [ML/T²]. (p. 415)" The kinematic (i.e., motions) equation is determined by the dynamic equation $x = A \sin(\omega t + \varphi)$, where A is the amplitude, φ is the phase, and ω is the angular frequency. "Since the amplitude and the phase are kinematic constants that depend only on initial conditions, they are arbitrary in respect to the dynamics. In contrast, the angular frequency is determined by the two dynamic parameters, $\omega = (k/m)^{1/2}$. (p. 415)" This ratio is dimensionally described as $([M/T^2]/[M])^{1/2}$, so the mass dimension cancels out and only the kinematic quantity [T⁻¹] is left. It is appropriate for the kinematic equation.

However, the kinematics used as information about the dynamics is determined by only the ratio of the dynamic parameters, k/m , not by specific values of either k or m . This causes the scaling problem of how the angular variable (i.e., only time dimensioned) can provide information about scale properties in events.

Subsequent studies have shown that temporal properties in events are uniquely linked to spatial properties, meaning that observers are able to judge scale properties (e.g., distance, size, mass, etc.) in dynamic events governed by gravity (Jokisch & Troje, 2003; McConnell, Muchisky & Bingham, 1998; Twardy & Bingham, 2002; Warren et al., 1987). Warren et al. (1987) found that observers were able to judge the elasticity of bouncing balls by detecting the single period duration when the height information was not available. McConnell et al. (1998) investigated whether observers were able to judge object size in event when only time and trajectory forms were available. They found that observers were able to use the information of time and trajectory to judge object size. Moreover, when observers were given the feedback on one event, they performed better and also generalized to other events as well as the one event in which the feedback was allowed. Thus, observers were sensitive to the general form of the spatio-temporal scaling relation, but feedback was needed to apply this relation to generalize events. Jokisch and Troje (2003) investigated whether observers were able to use the spatio-temporal relation to derive size information from point-light displays of dogs moving with varying stride frequencies. Animals as well as humans adjust their gait patterns in order to minimize energy consumption. For instance, the stride frequency of an animal is varied depending on its size for efficient energy consumption. The results showed that observers judged the size of dogs with high stride frequencies smaller than dogs with low stride frequencies. Thus, observers were able to use the physically determined relation between spatial and temporal scales to derive the size of a moving dog in the absence of other cues.

There has been more evidence that trajectory forms can provide information about an event so that dynamic events can be recognized. Bingham (1987) and Bingham et al. (1995) found that observers were able to discriminate inanimate dynamics (e.g., freely falling and bouncing objects or pendulums) from animate dynamics (e.g., a hand-moved spring or hand-moved pendulum). In comparison between a freely falling motion and one moved by hand, the difference between two events can be seen on the phase plane, where the kinematics were sampled directly from the video recordings.

Each event is produced by different dynamics that generated a trajectory form. The trajectory form is defined as the variation in velocity along a path of motion. For instance, a free fall and bounce produces a parabolic trajectory with a flat base corresponding to the impact, which exhibits characteristics of gravity. After the impact, the decrease of energy produces a decelerative parabolic trajectory.

In contrast, objects felled manually produce an elliptical trajectory with a half-flat base corresponding to inelastic impact and loss of energy, which exhibits characteristics of human limb movement. Then, an accelerative elliptical trajectory is shown due to energy increase. In these studies, results showed that observers were able to distinguish the inanimate motion of a falling and bouncing object from the animate motions produced by hand, even though those two objects were moved along the same path, to the same endpoints, and at the same frequency. Thus, it was shown that observers were sensitive to trajectory forms corresponding to dynamic events and used them to recognize events.

Wickelgren and Bingham (2001) found that 8-month old infants showed the ability to discriminate the differences in trajectory forms. They used three patch-light displays: a rolling ball, a water splash, and an occlusion event. The first and second events were asymmetric in time (i.e., nonreversible), whereas the last event was symmetric in time (i.e., reversible). In the first display, a ball started from one side of the screen, rolled across the screen and stopped at the opposite side. The second display was filmed looking directly down into a surface of water. The surface of some water covered by visible patches was splashed by dropping an invisible object, and then gradually settled in a damped oscillatory motion. The last display was a puppet swung out like an inverted pendulum from behind a wall. Researchers habituated the infants to the forward displays, and then switched to the reverse displays, or vice versa. It was found that infants habituated to the forward display, and then dishabituated to the reverse display or vice versa when nonreversible displays (i.e., a rolling ball and splash events) were used. In contrast, infants did not dishabituate when a reversible display (i.e., a puppet swinging out from behind the wall) was shown. The first two events looked totally different when shown in reverse, whereas the last event looked same even when shown in reverse. Thus, the results showed that infants also were sensitive to such variations in trajectory forms.

Muchisky and Bingham (2002) investigated the ability of adults to use trajectory forms as visual information about events. They tested whether observers were able to discriminate a freely swinging pendulum and a hand-moved pendulum by detecting trajectory form differences between two events. The freely swinging pendulum yields the symmetric trajectory form around the midpoint, whereas the hand-moved pendulum does not. They used a display of a circle oscillating in a straight path side to side on a computer screen.

In one study, they systematically varied the amplitude of the oscillating motion while preserving the period of oscillation. In the other study, they systematically varied the curvature of the elliptical trajectory form while preserving the symmetry of the form and the period and amplitude of the motion. After being shown a pair of displays, observers were asked to indicate which one was the hand-moved pendulum. The results showed that observers were able to detect small differences in trajectory forms. Observers distinguished an asymmetric trajectory form from the symmetric form and symmetric peaking or flattening of the form. These results are consistent with those of other studies, in which observers could detect properties of trajectory forms and use them as information to identify events.

IV. Conclusion

There has been a debate between the motor theory and the direct perception theory in speech and biological motion perception. The motor theorists claim that perception of speech and biological motions is special relative to perception of other events because only humans can both perceive and produce speech and biological motions. In other words, perception of speech and biological motions is distinct from that of other events because the two kinds of perception are processed by a different perceptual-motor system. Moreover, in biological motion perception, humans are more sensitive to their own motions relative to others' motions because they can use their own motor production (i.e., motor commands) as information to visually perceive self generated motions as well as speech.

In contrast to motor theorists, direct perception theorists claim that there is no distinction between speech and nonspeech perception or between biological motion and other object motion perception. As Gibson (1979) proposed, a pattern (e.g., optic and acoustic) in a causal source (e.g., light, sine waves, and motions) is used as information to perceive events.

In other words, every event is perceived directly by information from higher-order variables. For instance, different structures in sine waves specify different acoustic signals, and different structures in motions specify different dynamic motion events. In biological motion perception studies, it has been found that trajectory forms generated by dynamics of events can be used as information. Moreover, humans, including adults and infants, are sensitive to trajectory forms so that they can perceive and recognize events.

In sum, the motor theory claims that higher-order variables are decomposed by some kind of computation and reconstructed to perceive events, whereas the direct perception theory claims that higher-order variables are used directly as information to perceive events.

References

- Alexander, R. M. (1992). *The Human Machine*. NY: Columbia University Press.
- Beardworth, T., & Bukner, T. (1981). The ability to recognize one self from a video recording of one's movement without seeing one's body. *Bulletin of the Psychonomic Society*, 18, 19-22.
- Best, C. T., Studdert-Kennedy, M., Manuel, S., & Rubin-Spitz, J. (1989). Discovering phonetic coherence in acoustic patterns. *Perception & Psychophysics*, 45, 237-250.
- Bingham, G. P. (1987). Dynamical systems and event perception: A working paper, Part II. *Perceiving Acting Workshop Review*, 2, 4-7.
- Bingham, G. P. (1995). Dynamics and the problem of visual event recognition. In R. Port, & T. van Gelder (Eds.), *Mind as Motion: Dynamics, Behavior and Cognition* (pp. 403-448). Cambridge, MA: MIT Press.
- Bingham, G. P., Rosenblum, L. D., & Schmidt, R. C. (1995). Dynamics and the orientation of kinematic forms in visual event recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1473-1493.
- Bingham, G. P., & Wickelgren, E. A. (2008). Events and actions as dynamically molded spatiotemporal objects: A critique of the motor theory of biological motion perception. In T. F., Shipley & J. M. Zacks (Eds.), *Understanding events: From perception to action* (pp. 255-286). Oxford University Press, Inc.
- Cutting, J. E. (1981). Coding theory adapted to gait perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 71-87.
- Denes, P. B., & Pinson, E. N. (1996). *The speech chain: The physics and biology of spoken language*. New York: W. H. Freeman and Company.
- Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, 1, 121-144.
- Diehl, R. L., & Walsh, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of Acoustical Society of America*, 85, 2154-2164.

- Eimas, P. (1985). The equivalence of cues in the perception of speech by infants. *Infant Behavior and Development*, 8, 125-138.
- Eimas, P., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303-306.
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3-28.
- Fowler, C. A. (1989). Real objects of speech perception: A commentary on Diehl and Kluender. *Ecological Psychology*, 1, 145-160.
- Fowler, C. A. (1990). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of Acoustical Society of America*, 88, 1236-1249.
- Fowler, C. A., & Galantucci, B. (2005). The relation of speech perception and speech production. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (633-652). Oxford, UK: Blackwell.
- Fowler, C. A., & Rakerd, B. (1985). Work group on speech and sign language. In W. H. Warren & R. E. Shaw (Eds.), *Persistence and Change* (pp. 283-298). Hillsdale, NJ: Erlbaum.
- Gibson, J. J. (1970). On theories for visual space perception. A letter to Johansson. *Scandinavian Journal of Psychophysics*, 11, 75-79.
- Gibson, J. J. (1972). A theory of direct visual perception. In J. Royce & W. Rozeboom (Eds.), *The Psychology of Knowing* (pp. 215-227). New York and London: Gordon and Breach.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Jacobs, A., & Pinto, J. (2004). Experience, context, and the visual perception of human movement. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 822-835.
- Jacobs, A., & Shiffrar, M. (2005). Walking perception by walking observers. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 157-169.
- Johansson, G. (1970). On theories for visual space perception. A letter to Gibson. *Scandinavian Journal of Psychophysics*, 11, 67-74.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14, 201-211.
- Johansson, G. (1977). Studies on visual perception of locomotion. *Perception*, 6, 365-376.
- Johansson, G. (1985). About visual event perception. In W. H. Warren & R. E. Shaw (Eds.), *Persistence and Change* (pp. 29-54). Hillsdale, NJ: Erlbaum.
- Jokisch, D., & Troje, N. F. (2003). Biological motion as a cue for the perception of size. *Journal of Vision*, 3, 252-264.
- Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 812-832.
- Knoblich, G., & Prinz, W. (2001). Recognition of self-generated actions from kinematic displays of drawing. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 456-465.
- Latash, M. (1993). *Control of Human Movement*. Champaign, IL: Human Kinetics Press.
- Lieberman, A. M., Harris, K. S., Kinney, J., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, 61, 379-388.

- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 210-220.
- Mace, W. M. (1985). Johansson's approach to visual event perception – Gibson's perspective. In W. H. Warren & R. E. Shaw (Eds.), *Persistence and Change* (pp. 55-65). Hillsdale, NJ: Erlbaum.
- McConnell, D. S., Muchisky, M. M., & Bingham, G. P. (1998). The use of time and trajectory forms as visual information about spatial scale in events. *Perception & Psychophysics*, 67, 1175-1187.
- Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, 25, 457-465.
- Muchisky, M. M., & Bingham, G. P. (2002). Trajectory forms as a source of information about events. *Perception & Psychophysics*, 64, 15-31.
- Pisoni, D. B., Carrell, T. D., & Gans, S. J. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Perception & Psychophysics*, 34, 314-322.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, 212, 947-950.
- Runeson, S. (1977). *On the visual perception of dynamic events*. Uppsala, Sweden: University of Uppsala.
- Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, 112, 585-615.
- Stevens, J. A., Fonlupt, P., Shiffrar, M., & Decety, J. (2000). New aspects of motion perception: selective neural encoding of apparent human movements. *NeuroReport*, 11, 109-115.
- Todd, J. T. (1983). Perception of gait. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 31-42.
- Twardy, C. R., & Bingham, G. P. (2002). Causation, causal perception, and conservation laws. *Perception & Psychophysics*, 64, 956-968.
- Viviani, P., & Stucchi, N. (1992). Motor-perceptual interactions. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in Motor Behavior II* (pp. 229-248). Elsevier Science Publishers B.V.
- Warren, W. H., Kim, E. E., & Husney, R. (1987). The way the ball bounces: visual and auditory perception of elasticity and control of the bounce pass. *Perception*, 16, 309-336.
- Warren, W. H., & Shaw, R. E. (1985). Events and encounters as units of analysis for ecological psychology. In W. H. Warren & R. E. Shaw (Eds.), *Persistence and Change* (pp. 1-27). Hillsdale, NJ: Erlbaum.
- Wickelgren E. A., & Bingham, G. P. (2001). Infant sensitivity to trajectory forms. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 942-952.