MOTOR CONTROL: THE ROLE OF GOALS AND THE STRUCTURE OF 'MOTOR' REPRESENTATIONS

Silke Steininger
Institute of Phonetics and Speech Communication, Schellingstr. 3, 80799 Munich, Germany
kstein@phonetik.uni-muenchen.de

Jochen Müsseler
Max Planck Institute for Psychological Research, Amalienstr. 33, 80799 Munich, Germany
muesseler@mpipf-muenchen.mpg.de

ABSTRACT

A general model of motor control that is based on the following two assumptions will be presented: (1) Actions and perceptions are represented in a common domain (common coding principle, Prinz, 1997). (2) Actions are initiated by their effects or goals (principle of effect control). The model does not explicitly refer to articulation, but we assume that the principles we describe apply to speech as well. The two principles are derived from the '*T*heory of *E*vent *C*oding' (TEC; Hommel, Müsseler, Aschersleben, & Prinz, 1998). Up to now, this theory lacks concrete statements about how the action plans formed at the concept level are transformed into motor activity. We will outline how this gap may be filled.

Following a model of Kuperstein (1991) we assume that motor and sensory neuronal patterns are integrated into a topographic map. In this way, sensorimotor representations are formed that do not represent an event (a perception or an action) but the *common occurrence* of two events. To answer the question of how an association of a motor pattern with a subsequent sensory effect can be explained with a biologically plausible learning rule, we have modified a learning architecture from Aitken (1994). The resulting "motor" code consists of a number of sensory effects and a motor sequence that are associated in a special, structured way. An overview of empirical evidence for the model will be presented with a focus on perceptual–motor interactions.

1 INTRODUCTION

'We think the act and it is done' – with these words William James (1890, see Massaro, 1990) describes the assumption that voluntary acts are initiated by their anticipated sensory consequences. In the formulation of James this process seems rather simple; however, elaborated approaches have been developed to explain the initiation and control of actions. Traditional psychology, for example, regards motor control as a function that is separate from perceptual and other cognitive processes (e.g. Massaro, 1990). Nonetheless, in the area of Artificial Life and robotics and in current psychological theories there are notions that assume a close linkage between perception and action processes thoroughly in accordance with the formulation of James. They both stress the importance of the ability of a cognitive system to act and the importance of the fact that it is situated in an environment (AL: Pfeifer, 1996, Aitken, 1994; Robotics: Kuperstein, 1991; Psychology: Hommel et al., 1998; Hommel, 1997; for an overview: Steininger, 1999).

The present contribution is derived from this mutual basis. Whereas psychological approaches set action control at a higher cognitive level, AL–models refer mostly to the low level functions. We will show that both ideas can be combined.

2 COMMON REPRESENTATIONS FOR PERCEPTION AND ACTION

Two principles form the basis of our assumptions about motor control: The common coding principle and the principle of effect control. The common coding principle consists in the view that action and perception are not two separate systems, but fulfill two different functions that can be accomplished by the same representation. It somewhat resembles motor theories of perception (e.g. Liberman & Mattingly, 1985). The principle of effect control means that an action can be initiated by the (perceptual) goal of the action if the connection between the action effect and the motor program had been established earlier. This view is similar to the tasks dynamics approach (e.g. Saltzman & Munhall, 1989) which also stresses the importance of goal representations in motor control.

The model from which these two principles are derived – the 'Theory of Event Coding' (Hommel et al., 1998) assumes that the contents of perception and the goals of an action are both represented by codes of their distal features. Both perceived events (= perceptions) and events that are to be produced (= actions) are represented by an integrated network of *feature codes* that is formed as appropriate for a certain task. The features refer to kinesthetic or tactile effects of a motor movement *as well as* to distal effects of the movement like the light that goes on if a switch is operated. The integrated network is called an *event code*.

The forming of event codes is strongly dependent on the temporary goals of the perceiver/actor; this means they are tailored to the situational context. In TEC this is realized through the assumption of the variation of the relative weights of the feature codes that form an event code. Therefore if a feature is relevant for a task, it is weighted more strongly than the others (i.e., it can be activated more easily). In the case of perception, this weighting can be understood as an attentional process and in the case of action, as an intentional process.

3 THE ROLE OF GOALS FOR ACTION CONTROL

How do representations of events come into being? Hommel (1997) assumes that before an actor is able to produce goal directed voluntary actions, he passes through a 'motor babbling' phase: He produces uncoordinated, random movements that are triggered by external stimuli or by internal states. Next, an association emerges slowly between the produced effects and the activation pattern that caused the effects. A stable action representation is established when a certain movement is accompanied by the same effects, which strengthens the association between the movement– and

Proceedings of the 5th Seminar on Speech Production: Models and Data, Kloster Seeon, 2000, pp. 33-36.

effect–produced patterns.

A similar idea is used in the DIVA–model of speech production by Guenther (1995). He assumes that the babbling of infants comprises an action–perception cycle that can be used to tune the parameters of the production system. Through simulated babbling his model learns the mapping between orosensory (somatosensory) coordinates defining the shape of the vocal tract and articulator movements. In a second step, the mapping between sound targets and orosensory (i.e. vocal tract) targets is learned.

Another assumption related to TEC is the ideomotor theory of action control (e.g. Lotze, 1852). In contrast to early versions of this theory, however, *all* effects are assumed to be connected to the motor activation, not only body–related afferent effects (proprioceptive or kinesthetic effects) but also distal sensory effects (such as the corresponding lighting of a lamp). When motor patterns occur repeatedly with corresponding distal effects, random effects become rarer and rarer and an increasingly stable representation of the action plan is formed. However, TEC (Hommel et al., 1998) makes less concrete statements about how the action plans (formed at the concept level) are transformed into motor activity. In combining the views of TEC with an AL model of motor control this gap may be filled.

The AL–learning architecture from Aitken (1994) allows a mobile agent to acquire sensorimotor representations. Aitken expands ideas of Kuperstein in assuming a network of target maps that are connected together (the socalled '*S*ensory–*M*otor–*A*ssociation–Modules', in short SAM–modules). All modules function by the same principle: If all or a subset of their inputs are correlated (i.e., if they are repeatedly active together), this correlation is stored in the module.

How the ability of the module to detect and to store the correlation is realized, is not important here. One can assume a neuronal network or a symbolic system of classification rules. It only has to be guaranteed that correlations can be detected and stored through unsupervised learning. Furthermore, random input should be ignored and noise should be tolerated.

The learning principle according to Aitken (1994) is based on the fact that for every input–vector a distribution of possible output states is formed (as a function of the input states). By means of this distribution, the output state is then selected. After that, the distribution is actualized in such a way that the correlated inputs later evoke the same output state with a greater probability. Uncorrelated input therefore leads to random output and correlated input after the learning always produces the same output. Competition for representational space ensures that correlations subsequently unrepresented in the inputs are eventually replaced by stronger correlations.

In order for the module to learn sequences of movements, Aitken assumes that two of these modules are connected in a special way. The so–called sensory module gets input from the environment and from the motor module and passes on output itself to the motor module. The so–called motor module gets input from the sensory module and passes on output to the environment and to the sensory module. The output of the motor module to the environment produces effects there. For the SAM–modules learning is possible because the inputs to the sensory module are no longer random because of the connection of the sensory module and the motor module to the environment. Motor output has effects in the environment which leads to non–random input sequences at the sensory module. Because the modules can detect correlated inputs, they can learn to link the state of the motor module with the state of the sensory module. How does this work in detail?

The first step consists in 'motor babbling' as described above. The random output of the motor module produces random movements. After some time a certain motor sequence will occur repeatedly and will lead to effects that can be detected. The repetition can be caused by anatomic structures that suggest a certain repetition or by realities in the environment that lead to a repeated performance of the same movement sequence. This allows the sensory module to detect the correlation between the input from the environment (the *effect of a former state* of the motor module) and the *current state* of the motor module.

After learning has taken place, the sensory module has representations of the correlated occurrence of a motor state and the sensory effect that was caused by the motor state one time–step before. In principle, a motor command sequence with its sensory consequences is learned (what Aitken calls a 'primitive effective sequence'). This is the first component of learning.

The second component of learning takes place in the motor module. Because the output state of the motor module is connected through its effects to the input of the sensory module, certain motor states occur consistently with representational states of the sensory module. If the motor module is in state C, the sensory module is consistently in state [B,E(A)] (i.e., B correlated with the effect of A). If a certain input occurs, for example B[E(A)], the motor module will produce – because of its learning rule – the appropriate output with a higher probability in the future (i.e., C), when both occur together repeatedly. This means the motor module learns the appropriate motor answer to a state of the sensory module. These two steps allow the learning of a motor sequence that occurs repeatedly during motor babbling.

4 THE STRUCTURE OF "MOTOR"–REPRESENTATIONS

To find a solution to the problem of how a certain motor pattern can be connected with its respective *subsequent* effect, we assume that the sensorimotor code consists of an association of a motor sequence that was learned in the way described by Aitken with several sensory effects. What does this mean in detail?

We assume that the common coding domain is structured similarly to the topographic map in the model of Kuperstein. The resulting codes possess distal reference as demanded, because only regularities in the environment can lead to a correlation of simultaneously active sensory and motor patterns. The result are correlation codes that represent a relation between a sensory input and a motor activation.

As the model of Aitken makes clear, some problems have to be solved to explain the acquisition of event codes: To enable effect control of actions, a correlation code has to make a connection between the motor activation and its subsequent sensory effect, that is, the code has to have a *directional temporal structure*. Furthermore, the motor activation has to be linked to its *proper* effect and not to the effect of the previous motor activation as is the case in the model of Aitken. And last, but not least, there remains the problem of how a correlation between an input and an output can be formed: To combine a sensory and a motor pattern with his target map, as proposed by Kuperstein (1991), both must be available as input. The goal, however, is to trigger the appropriate motor pattern by the sensory effect; therefore it is not enough to combine a sensory and a motor representation. The motor representation actually has to make sure that the muscles are activated, that is, it has to produce output. Answers to all three problems can be found when elaborating on the ideas of Aitken (1994):

The last problem, the manner in which the motor output can be used as input, is solved by Aitken through the assumption that his motor module passes on output both to the environment and to the sensory module. In the sensory module then, the sensory input (an effect of the motor output to the environment) can be correlated with the motor output. This conforms exactly to the assumptions of Hommel (1997) in regard to his assumption about the acquisition of the effect control. However, the combination of motor activation and its effect (= sensory input) always occurs with a temporal delay; an effect occurs after its trigger. In Aitken's module, therefore, the input of an effect is correlated with the motor state that *follows* the effect–triggering motor state. This answers the question of how sequences of effect–movement series can be learned, but it does not answer the question of how the system can learn the connection between a certain motor output and the *appropriate* effect or input; only simultaneous events are correlated.

As a solution to the two first problems we assume that a correlation code consists of the integration of the 'primitive effective sequence' described by Aitken. It is assumed that the code is structured like the sensory and the motor module of Aitken, by which the learning of a primitive effective sequence is possible. This whole sequence then is 'integrated', meaning it becomes possible to activate it as a whole without losing its basic organization. The code in principle consists of a 'temporal averaging' over the individual effects of a primitive effective sequence that stay associated with the sequence as a whole. We then assume that each sensory input that conforms to an individual effect of the learned movement sequence activates the whole code and leads to the triggering of the whole movement. Because of the underlying temporal organization that consists in the correlations between the sensory and the motor representations of the code, it is assured that the individual motor activations take place in the correct order.

So, the presented correlation code links several effects with a corresponding motor pattern or, in other words, with an elementary (mini–) movement sequence. The effects are coded as simultaneous because of the temporal averaging, however the underlying correlational structure of the learned sequence guarantees a directional temporal organization of the code. The three above mentioned problems are solved by these assumptions.

In our opinion, event codes develop from sensorimotor correlation codes that represent an elementary mini–movement sequence. The sequence is associated with several sensory inputs/effects. The correlation codes are presumably acquired through 'motor babbling' (see above; see also Hommel, 1997; Guenther, 1995; Kuperstein, 1991). Movements are restricted by the anatomy of the learning system and the environmental conditions, just as analogously the sensory information is restricted by the environmental conditions and the structure of the sensory organs. The correlated occurrence of sensory and motor information therefore is determined by the structure of the environment and the structure of the body of a newborn or the structure of the effectors of an artificial cognitive system. The resulting codes then represent mini–units of this structure of the environment and the anatomy or, to be precise, their interaction. The correlation codes that form the basis of the system therefore depend heavily on its physical structure and the set–up of the environment. This makes clear why it is important to investigate perception and action together and with regard to the body in which cognition is 'implemented' (see Pfeifer, 1996).

The described correlation codes could be called *subfeature codes* in contrast to the event codes of TEC. They too are event codes with distal reference, but they refer to mini–events or implicit knowledge. They possess neither the abstraction level of the event codes described in TEC nor that of the feature codes. With the name 'subfeature codes' we want to make clear that the codes are not the parts of a symbol (in the sense of symbolic information processing), but the parts of a connectionistic representation. The second important property of these representations is that they do not map an event in the environment but a *relation* between an event and the 'owner' of the representation. They represent regularities in the interaction between the actor/perceiver and his environment. Common coding representations of the described kind get their reference by their history of origins; they are automatically anchored in the world.

5 EMPIRICAL EVIDENCE

The assumption of a close connection between perceptual processes and processes of action control is supported by empirical data. These data show an interaction of both functions if the similarity of action and perception tasks is varied. For example, compatibility experiments aim to investigate how far a stimulus that shares features with a reaction is able to facilitate or handicap this reaction. The general finding in such experiments is that reaction times are faster with compatible stimulus–response–mappings, for example, when the spatial layout of the responding keys corresponds to the spatial layout of the stimuli. In contrast, reaction times are slower with incompatible stimulus–response–mappings (for an overview, see Alluisi & Warm, 1990). This is generally explained by the assumption of feature overlap ('dimensional overlap', e.g., Kornblum, Hasbrouq, & Osman, 1990), where a compatible layout primes the correct answer automatically. With an incompatible layout, the wrong answer is assumed to be activated first and the subject reacts slower because it takes time to cancel this activation. Without the assumption of common codes for perception and action, it is hard to understand why these privileged relations between stimuli and reactions exist (Prinz, 1997).

If action and perception share the same representations, this should not only lead to mutual priming. It also could lead to a conflict if both functions have to use the same resources at the same time. Müsseler and Hommel (1997) were able to show that a perceptual impairment can be observed when an action is produced simultaneously. The experimental setup was like this: In every trial the subjects had to accomplish two tasks that were connected in a special way. The first task consisted in the execution of an obligatory pressing of both buttons of a computer mouse and then a directly following left or right button press. The last, single button press was predetermined at the beginning of a trial by an imperative stimulus. The second task consisted in a perception task. The double mouse click triggered the short presentation of an arrow head on the display that was masked by a random dot pattern. The subjects had to identify this arrow head. The presentation time of the arrow head was so brief that the subjects made 10–40% identification errors on average.

The main point in this experiment was the combination of the two tasks: The double button press triggered the presentation of the stimulus to be identified so that the stimulus appeared at a point in time when the single button press was being prepared. In several experiments of this kind, a decrease in identification performance of 10% to 20% was found in the compatible condition. In other words, a left arrow could be identified worse in comparison to a right arrow if simultaneously a left action was executed and vice versa. An action that shares features with a simultaneously to be identified stimulus therefore disturbs the perception of this stimulus. Again, this finding is hard to explain without assuming common representations for perception and action.

In further variations of this experiment we investigated to what extent it is the effect of the action that influences the stimulus identification (Steininger, 1999). For example, the subjects were required to evoke right and left arrow heads on the display, however not with right and left button presses but with single and double button presses. A feature overlap existed only between the visual action effects and the arrow heads to be identified. Under these conditions a decrease in detection performance was also observed; however, it was not so pronounced as in the original experiments; presumably, because of a smaller feature overlap between stimulus and response.

6 CONCLUSION

Our ideas have led to a model that describes how the general assumption of effect control of actions in artificial or human cognitive systems can be realized. To what extent the described sensorimotor correlation code is actually used in human information processing remains an open question. Nonetheless, this assumption presents a basis for further investigations in the light of the empirical evidence supporting it. Another advantage is that the model does not rely on complicated, biologically implausible learning algorithms (only correlations similar to learning à la Hebb are necessary). In respect to speech perception and production, we assume comparable underlying processing principles, because speech can be considered a special case of motor control. Therefore, one could speculate that the representations for speech control consist of small learned motor sequences of the articulators that are associated with the kinesthetic and/or acoustic (or even visual) effects of the movement. Thus, the present framework also seems to be a promising starting point for further investigation in the area of speech.

REFERENCES

Aitken, A. M. (1994). An architecture for learning to behave. In D. Cliff, P. Husbands, J.–A. Meyer & S. W. Wilson (Eds.), From animals to animats 3. Proceedings of the Third International Conference on Simulation of Adaptive Behavior (p. 315–324). Cambridge, Mass.: MIT Press, Bradford Books. James, W. (1890). Psychology (Briefer Course). New York: Holt.

Alluisi, E. A. & Warm, J. S. (1990). Things that go together: A review of stimulus–response compatibility and related effects. In R. W. Proctor & T. G. Reeve (Eds.), Stimulus–Response–Compatibility (p. 3–30). North–Holland: Elsevier.

Guenther, F. H. (1995). Speech Sound Acquisition, Coarticulation, and Rate Effects in a Neural Network Model of Speech Production. Psychological Review, 102 (3), 594–621.

Hommel, B. (1997). Toward an action–concept model of stimulus–response compatibility. In B. Hommel & W. Prinz (Eds.), Theoretical issues in stimulus–response compatibility (p. 282–320). Amsterdam: North–Holland.

Hommel, B., Müsseler, M., Aschersleben, G. & Prinz, W. (1998). The theory of event coding (TEC). A framework for perception and action. Max Planck Institut für Psychologische Forschung, Abteilung Kognition und Handlung, Paper No. 3/1998. (submitted)

James, W. (1890/1981). Will. In James, W. (Ed.), The principles of psychology (Volume II, p. 1098–1193). Cambridge, Massachusetts: Harvard University Press.

Kornblum, S., Hasbrouq, T. & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus–response compatibility – A model and taxonomy. Psychological Review, 97, 253–270.

Kuperstein, M. (1991). INFANT neural controller for adaptive sensory – motor coordination. Neural Networks, 4, 131–145.

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revisited. Cognition, 21, 1–36.

Lotze, R. H. (1852). Medicinische Psychologie oder die Physiologie der Seele. Leipzig: Weidmann'sche Buchhandlung.

Massaro, D. (1990). An information–processing analysis of perception and action. In O. Neumann & W. Prinz (Eds.), Relationships between perception and action: Current approaches (p. 133–166). Berlin, New York: Springer.

Müsseler, J. & Hommel, B. (1997). Blindness to response–compatible stimuli. Journal of Experimental Psychology: Human Perception and Performance, 23, 861–872.

Pfeifer, R. (1996). Building "fungus eaters": Design principles of autonomous agents. In P. Maes, M. J. Mataric, J.–A. Meyer, J. Pollack & S. W. Wilson (Eds.), From animals to animats 4. Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior. Cambridge, Mass.: MIT Press, Bradford Books.

Prinz, W. (1997). Perception and action planning. European Journal of Cognitive Psychology, 9 (2), 129–154.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. Ecological Psychology, 1, 333–382.

Steininger, S. (1999). Wahrnehmung und Handlung. Eine experimentelle Analyse einer Wahrnehmungsbeeinträchtigung bei simultan ausgeführten Handlungen. Aachen: Shaker.