Working Models of Perception; Five General Issues

Willem J.M. Levelt*

1 Introduction

There are, at least, two kinds of scientific meetings. The first, homogeneous, kind brings together people from the same discipline who work on closely related problems, and hope to jointly clarify one or two of them. The second, heterogeneous, kind collects scientists from different backgrounds on a fairly global theme, and the intention is one of mutual instruction and general enrichment. Both types of meetings have their specific problems. The homogeneous meeting, often called a 'workshop', carries the risk of being utterly boring, because all participants already know beforehand the songs of most of their colleagues. The heterogeneous meeting, often called a 'symposium' in memory of Greek drinking-bouts where one indulged in more or less decent talk and rhetoric, risks being too difficult or too superficial. It is often hard to follow a serious detailed paper from another discipline; when this leads to making major concessions, one may end up in the shallow table talk which lent its name to this kind of meeting.

Our present symposium was clearly tilted to being of this latter heterogeneous type. A collection of physicists, engineers, psychologists, computer scientists, linguists, logicians and phoneticians addressed issues of human perception from their often wildly different perspectives. As Dutch idiom has it, we have been balancing on the sharp of the edge. We were, on the one hand, treated to glimpses of exciting new developments in basic and applied perception research. Time and again I regretted not being more of an insider, so as to be able to share the full excitement of discovery and argumentation. And indeed there were occasional moments where I had not the faintest idea what was at issue. But by and large the excitement surpassed the frustration. On the other hand, we have had our moments of table talk. Superficial maybe, but not less pleasant. It contributed to creating the comfortable illusion that human perception is one, and that we

* Max-Planck Institut für Psycholinguistik, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands.

WORKING MODELS OF HUMAN PERCEPTION
ISBN 0-12-238050-9

have all been chosen to participate in its mystery.

My present task is, I fear, to contribute some more to this latter illusion. I am supposed to elaborate where we are one already, and to highlight some of the remaining obstacles on the twisting road to unity. In short, I must address some general issues in perception. I have accepted this task in full awareness of threatening shallowness; but the inevitable liturgy of rounding up a conference will hopefully raise your tolerance level.

I intend to address the following five general issues. The first one is: How do models of perception handle convergence in real time? The second one is: How, in models of perception, are structure and function determining of process? The third issue is: What is the learnability status of models of perception? The fourth one is: What is the role of input management in perception? And the last issue is: What is general and what is specific in perceptual architecture? The exercise will mainly consist of raising and elaborating the questions. I felt that, even after this successful meeting, we should leave with the feeling that there are still some questions to be answered.

2 Issue 1: How do models of perception handle convergence in real time?

The ability to converge is a litmus test for any model of perception. A general paraphrase of this ability is: Given context and well-formed input, the model should converge on the ecologically correct solution. And if it does, the mediating processes should be able to run in real time on the neural issues involved. If a model fails to live up to these criteria there is something seriously lacking. Elsewhere (Levelt and Flores d'Arcais, 1987; Levelt and Schriefers, 1987) I have argued that all existing models of lexical access, including logogen theory and cohort theory, fail on this criterion. They cannot guarantee that, given the stimulus, the correct word identification is going to be made. Cohen, in his introductory lecture, acknowledged the dawning awareness of this problem among speech researchers when he reviewed the change of terminology from optimistic 'speech analysis' via 'speech recognition' and 'speech understanding' to all-encompassing 'speech interpretation'.

Probably the most explicit effort to implement the convergence criterion we have seen during this conference was Rosenbloom's SOAR model. SOAR is built in such a way that, *if* it comes up with a solution, it comes up with

the correct solution. The learning errors that it still makes occasionally are to be *avoided* in better versions of the model. The approach is extremely conservative, in that the power of SOAR increases from version to version, without ever relaxing this correctness criterion. At the same time, the authors are highly aware of the real-time issue. It may not be so hard to come up with the correct solution given sheer infinite processing time, but it becomes very hard to approach the speed of human processing without blocking or coming up with an erroneous result. Two features in the model are material in approaching realistic temporal parameters. The first one is parallel processing, the simultaneous elaboration of different goals on a stack. The second one, Rosenbloom's warranty of quality, is chunking. Previously elaborate, but successful trains of productions are chunked together to be available for application at one swoop. It is precisely this feature which makes SOAR relevant for the study of perception. Most definitions of perception acknowledge its character of immediacy, as opposed to problem solving where there is often awareness of effortful steps and intermediary results. But skilled problem solving does have perceptual features, as De Groot (1946) argued long ago when he discussed the skillful perception of configurations in chess. It is this kind of high-speed perceptual process which gets modelled by chunking. And Bösser, in his reaction to Rosenbloom's paper, was right in stressing that this should be an obvious extension of SOAR.

One of the most serious problems in meeting the criterion of convergence is the interaction of stimulus and context, and for which Cohen reintroduced Wundt's good old term *apperception*. We have seen several cases of this during the conference. Nooteboom, for one, presented the case of voice quality, which is – in part – determined by the correct perceptual assignment of perceptual noise to voice. But if the speech is not coded in the right way, the perceptual system throws out that noise as annoying 'context', instead of taking it as a feature of voice quality. Anstis showed how the same visual stimulus could be interpreted one way or another, dependent on various kinds of context. He demonstrated convincingly that a model of motion perception will not uniquely converge, given a stimulus, without specifying such contextual conditions as the gradients of luminance change and the amount of surrounding dynamic noise. Juola correctly stressed that the global context effects demonstrated by Anstis cannot be explained by the existing models. Something extra is needed. Houtsma suggested that this extra is probably quite central, because the ear shows the same type of context effects. This was an elegant non-sequitur, quite relevant for the fifth issue discussed in this paper. In some of his published articles,

Anstis has shown that motion interpretation also depends on the pattern of neighbouring motions, and even on the interpretation of the patterns involved as real objects. The latter finding is particularly upsetting, because the pattern's constitution as an object may itself be a consequence of the induced apparent motion. A working model should tell us what is chicken and what is egg.

Leeuwenberg approached the context problem in a totally different and quite principled way. He declared visual context to be part of the stimulus. The simplicity of coding the whole lot determines the percept. The notion of 'likely interpretation' is no more than a derived one. It seems to me that Leeuwenberg's coding theory may in this way handle some of Anstis' published cases correctly, i.e. by just using the precisely defined simplicity criterion, convergence will be perfect. On this view, a perceptual system is a purely syntactical machine which delivers its solution irrespective of world knowledge, circumstantial likelihood, or what have you.

I am still very sympathetic to this view, in spite of Sutherland's eloquent reintroduction of the likelihood principle. Of course, he was right to indicate that simplicity should always be seen in relation to a notation system. But Leeuwenberg's notation system is by no means an arbitrary one. It was precisely designed to capture major regularities in perceptual encoding. The view of the perceptual system as a syntactic machine is akin to Fodor's modularity notion: Input systems are reflex-like working modules, whose functioning cannot be affected by other processors. On that view, they are in particular impenetrable to the perceiver's convictions, beliefs, or intentions. The main problem with this notion is to come up with the correct definition of the system's output. What kind or level of representation is computed by this modular syntactic machine? When I see a face and immediately recognize it as Herman Bouma's, what was the module's part in this? Was its output a bunch of overlapping and connected surfaces with their textures and colors? Or was it already a 3-D interpretation? Or was it a mapping onto the perceiver's stored Herman-template? The latter Leeuwenberg surely wouldn't like. In fact, it has been suggested that the visual module's output is a kind of 2.5-D representation in Marr's sense. This comes reasonably close to Leeuwenberg's coding, although the theory is very different in other respects.

Assuming that such questions can be answered, the convergence problem takes a different shape. Phenomenologically speaking, the perceiver converges on 'Herman's face', not on some uninterpreted 2.5-D representation. It is not at all despicable for a perceptuologist to try and account for

phenomenological convergence, the perceiver's own full interpretation of the scene. Sutherland correctly stressed that the task for visual perception is to recognize objects, with their shape, size, and distance properties. When it is assumed, for the moment, that some version of modularity is correct, there are then two tasks to be faced. The first one is to handle convergence in the module, which I will call *modular convergence*. An algorithm which automatically derives a unique and correct Leeuwenberg coding from a scene would do a thing like that. The second one is to map that intermediary representation onto the perceiver's final full interpretation, i.e. something like 'Herman's face'. This I will call *post-modular convergence*. This second step may also be highly automatic, but it is not modular. Expectations, beliefs, desires, intentions may have their share in the interpretative process. It is here that likelihood does matter.

The first kind of 'modular' convergence may be hard to attain for a model of perception, but the latter kind of 'post-modular' convergence is even more horrendous. This is because the system leaks like a sieve. It is at the mercy of factors such as analogy, recent experience, suggestion, emotion, and the motivational system Bösser referred to in his reaction to Rosenbloom. It should immediately be added that many do not share the distinction between a modular input system, connected to non-modular more central interpretative processors. Connectionists are of this kind, and so are Gibsonians. Both are optimists as far as convergence is concerned. Connectionists live in the paradisical belief that convergence will naturally fall out if a simple non-specific learning algorithm is combined with some finite sequence of activation of 'learning' pairs. Gibsonians tell us that the perceptual system has very little to do because the invariance is in the stimulus. This enables a one-to-one mapping to the correct interpretation. There is no convergence problem.

But alas, both views are naive. I will not return to the connectionists' beliefs, since, honestly to my surprise after Cohen's (somewhat ambivalent) insistence that we consider connectionism, nobody defended those beliefs during this meeting. Lindblom did a professional job on the Gibsonians. He discussed the hopeless non-convergence of speech perception models. And he pointed to the inherent weakness of theories which tacitly or overtly assume that the stimulus does contain the relevant information. His example stimulus [lesnsevn] is just as ambiguous as some of Anstis' apparent motion stimuli, or the anaphoras Bunt talked about. Only context can disambiguate them. If the perceptual model does not incorporate a means of letting relevant context have its effect, the model simply fails on the convergence

criterion. But the comparison with Anstis' cases of visual ambiguity is instructive. There, following Leeuwenberg's lead, one could conjecture that disambiguation or convergence can be handled entirely within the modular stage of the model, by taking context to be part of the stimulus. Lindblom's case, however, resists such an approach. The ambiguous stimulus is all there is to be heard. The ecologically valid interpretation only results when knowledge of the discourse situation is taken into account. Convergence must be post-modular, if we stick to that distinction. At any rate, the disambiguating information is not part of the immediate perceptual scene.

If the ultimate goal is the ecological validity of a perceptual model, i.e. correct convergence in real-life perceptual situations, it is advisable to insert a phenomenological stage early in one's research endeavours. Lindblom announced doing just that, collecting natural spontaneous speech to find out what the ecologically normal stimulus looks like, as opposed to the perfected laboratory case. Bunt also based his model of text processing on elaborate analyses of natural dialogues. The gain over standard models of text interpretation is enormous, just because these analyses made him aware of the necessity to build a powerful pragmatic component into the model. There will be no correctly converging text interpretation without letting beliefs and intentions have their way.

But let us not become self-congratulatory too quickly. It is one thing to be explicitly aware of the context- and knowledge-dependency of convergence, it is quite another thing to create a working model that handles it correctly, or at least major aspects of it. Lindblom, for instance, has no working model which guarantees convergence. Bunt's model can surely be fooled, as can most other models that we have reviewed these days. This is in no way tragic, it just shows that we need research funds for some more years to go.

3 Issue 2: How, in models of perception, are structure and function determining of process?

Structuralism and functionalism are ill-defined notions, but it is worth considering the following two approaches to perception. One is to dissect the structure of the system (be it anatomically, physiologically, psychologically), and predict from there what the system will do, what kind of representation it will generate. Let us call structuralism the approach of predicting process from architecture. The other is to analyse what goal is served or function is

performed by the system, and then to conjecture an architecture which can compute that function in real time. Let us call functionalism the approach of predicting processing architecture from function. These approaches are usually intermingled in our scientific minds, and that is how it should be. The issue becomes more serious, however, if causative claims are being made. Do structures necessarily cause particular perceptual results? Or does the processing architecture necessarily adapt to the functional requirements?

Anstis came close to the causative structuralist position when he predicted that segregation of motion *should* be a consequence of the existence of a neural channel for luminance change. The prediction did not bear out. O'Regan similarly derived *necessary* perceptual constraints from the grain of the retina. But Taylor and Houtsma made clear that structure does not necessarily restrict functional resolution that way; it need not in a low-noise system, for instance. Also Goldstein came close to the causative structuralist position when he suggested that the non-linear nature of cochlear filtering should quite likely reflect itself in speech perception. The structural cause is there, hence the perceptual effects are to be found. It is interesting to compare this to Atal's finding that even rather heavy distortions in the speech signal, namely in fixed multi-pulse coded speech, have negligible perceptual effects. Another such case presented during the conference by Taylor is Bouma's systematic distortion of letter forms in handwriting, which apparently hardly interferes with word recognition. If such heavy distortions in the stimulus go unnoticed, why should not non-linear distortions created in the peripheral sense organs go unnoticed? Though this may all be obvious to functionalists, they have no principled way of predicting which kind of sensory distortions will or will not have perceptual effects. The structuralist predictions are at least *predictions* worth testing. They direct our attention to potentially interesting phenomena.

Functionalist positions were taken rather more clearly by Rosenbloom and by Lindblom. Rosenbloom, working in a Newellian framework, has the architecture of his model shaped by the task and by past experience with the task. Lindblom introduced the notion of 'adaptive variability', the adaptation of both production and perception of speech to current functional demands – in fact a frontal attack on the mainstream structuralist practice in linguistics.

Lindblom did not hesitate to extend this notion to the explanation of the cultural history of languages. Different languages are different solutions to the joint optimization of motor effort and perceptual distinctiveness. This is, in my view, a highly attractive functionalist notion. Still, not everybody

will be happy with it. Ohala, for instance, was not because there is no real *improvement* in language evolution. I also suspect, for instance, that it makes Leeuwenberg shudder. He senses circularity in such evolutionary explanations, as appeared from his discussion with Sutherland about the phylogenetic likelihood of perceptual schemes. According to Leeuwenberg's structuralist point of view, the perceptuologist's task is to find out what kind of representation the system computes, and according to what principles. The statement that this is in some way or another a 'best' solution cannot be falsified. The functionalists, I feel, have a strong point. But they easily force themselves into a position where they can only test the null-hypothesis.

The issue of functionalism is closely connected to the issue of learnability. Learning is a form of optimization with respect to a certain task or function. This brings us to issue 3.

4 Issue 3: What is the learnability status of models of perception?

Another litmus test for the adequacy of a perceptual model is that it can be acquired in finite time. Some perceptual skills are preprogrammed, being ready at birth, or maturing shortly after. Other skills require substantial learning. Making certain categorical phonetic distinctions, for instance, is to a surprising degree already there right after birth; correctly recognizing the words of a language takes years to develop. A working model of perception must be learnable in realistic time. By this I mean that there must be a learning algorithm which, given the prewired capacities, the perceptual input, and the feedback, produces the working model as output, and this algorithm should be consonant with speed of acquisition in the human learner. The initial learning algorithm can, of course, only be prewired.

It is still quite unusual for students of perception to bother about learnability. This in spite of the fact that learning plays a crucial role in perception, as we all agreed during the Sutherland/Leeuwenberg discussion. The issue is of great importance for two reasons. The first one is that an unlearnable model is ipso facto unacceptable. If one can show that there is no way for the organism to acquire the supposed processing architecture, then that model must be wrong. Take as an example Leeuwenberg's structural information theory. It can be considered as a grammar which assigns structural descriptions to perceptual patterns. We do not know whether that grammar is learnable. Its learnability will, in part, depend on what we assume to be

innately given as abstract coding schemes. It also depends on the kind of feedback we suppose the perceiver receives in the course of acquiring the grammar. And its learnability will, last but not least, depend on the generative power of the grammar. Can we characterize the set of patterns that are well-formed in terms of the grammar? A grammar's power is, in fact, the main determinant of its learnability. A similar question can be asked with respect to Groenendijk's Dynamic Logic. If it is indeed the appropriate model for the semantics of conversation, than it should be learnable in finite time.

The second reason to be concerned with learnability lies in the *course* of perceptual learning; for both theoretical and applied reasons it is important to know *how* a perceptual skill is acquired over time. It is one thing to know that acquisition of a model is feasible, it is quite another thing to design the learning algorithm. Rosenbloom practically reversed the order of dealing with these issues. He designed his chunking as a learning algorithm. The resulting processing architecture is then, ipso facto, a learnable one.

That approach will seldom be taken by students of human vision or audition. They first create the model, and only much later, if at all, wonder about perceptual learning. The situation is, however, quite different in the applied fields. Engineers have long known that image coding can be enormously improved by implementing an acquisition procedure in the system, or so I thought. This is relatively easy if the set of patterns to be recognized is of a restricted kind, such as the letters of the alphabet, as used in optical readers (e.g., in Kurzweil machines). But also automatic speech recognition systems that learn, involving far larger sets of stimuli, are now showing their market value, and generally the value of learning algorithms in technological applications. I was surprised by Kunt's review of image coding where no reference was made to learning algorithms. And also Mussman did not mention anything of the sort. Do I overestimate the currency or relevance of artificial learning systems? I probably do.

That first morning session on image coding surprised me for other – but still related – reasons as well. How can an important field of technology so substantially ignore existing theories of visual perception? Following up on a private exposé Goldstein gave me later that day, we can make a distinction between waveform encoding and source encoding. One could be reasonably happy with the presented work as far as waveform encoding is concerned. One only needs Fechner's law, delivered over a century ago, or something similar. For source encoding, however, one needs a *model.* For speech encoding the classical model is LPC, a model of the speech producing apparatus.

Atal showed how this was not good enough, and how it became replaced by increasingly sophisticated perception-based models, such as CELP. For the encoding of visual images, one also needs some perception-based model. But as long as this necessity is virtually ignored, the coding will be hopelessly sub-optimal, as was implied by Taylor's questions during the discussion. Atal said at one moment *"speech coding and auditory research go hand in hand"*. One should wish that a few years from now a *mutatis mutandis* statement can be made for visual pattern coding and visual research. In addition – and this brings us back to the present issue – the question of learnability cannot be professionally addressed without a model, since what is learned is precisely the model.

Other applied fields where theories of perceptual learning are indispensible are the education of the handicapped, and human-computer interfacing. Fourcin reminded us of the learning problems of the hard of hearing when they want to acquire the phonetic contrasts of their language. The learnability of these contrasts by hearing-impaired children can be understood from the auditory pattern model. It says that all major phonetic contrasts of a language are associated with *families* of acoustic/auditory features. The absence or attenuation of one feature can be compensated for by the presence of another feature. One might say that the family of features forms an *equivalence class* with respect to a particular phonetic contrast. With adequate training, hearing-impaired children can acquire a contrast by using acoustic-auditory features that are still available to them.

During the discussion of this paper, Houtsma referred to cross-modal training programs, such as the one developed by Povel in Nijmegen, where equivalence classes of acoustic features are factored out from the signal and mapped onto some visual dimension, a different but 'natural' one for each phonetic contrast.

The importance of models of acquisition for human-computer interfacing needs no extensive arguments either. In fact, this kind of learning has lately become a pretty active field of research. User-friendliness is almost synonymous with easy learnability. Here again, a main issue for learnability, as Wright pointed out, is the 'natural' factorization of design features. 'Natural' being independently recognizable, 'chunkable', and learnable by the user.

5 Issue 4: The management of input

When I discussed convergence as a litmus test for the adequacy of a working model of perception, I distinguished between 'modular' and 'post-modular' convergence, where modular output is input to a component which can acknowledge context, expectations, motivations, and so forth. But the picture has to be complicated even more. In many, if not most perceptual situations input is made dependent on expectation, motivation, earlier interpretation, etc. Let us first consider the case where this is controlled by the perceiver himself. One major example at this symposium was provided by O'Regan. Eye fixation is selection of input, and fixation is by no means a random process. Rather, it is in some sense based on expectation. O'Regan could show that the saccades in reading are programmed in such a way that the resulting fixation approaches the optimal or convenient viewing position for a word, i.e. the position allowing for speediest convergence in word recognition. The expected most convenient position is derived from peripheral vision during the previous fixation. Peripheral vision does not allow for the computation of the *real* most convenient viewing condition, because that requires, of course, recognition of the word. But, given reading experience in the language, a fair guess can be made on the basis of physical properties of the word. In the experienced reader this turns into a purely ocular-motor response. Taylor suggested that that preferred viewing position is systematically deviant from the convenient viewing position. O'Regan could handle these discrepancies elegantly. At any rate, if fixation turns out to be too far from optimal, local tactics may correct it. There is, in short, an elaborate system of input management, optimizing the chances of correct convergence, whose crucial component is a feedback loop.

As we all know, not all feedback loops in perception are of a post-modular kind, i.e. acknowledging the perceiver's expectations. The pupil is automatically adjusted to the intensity of the stimulus, and the eye automatically follows a moving object. For audition, similar short-circuited feedback loops are involved, attenuating the input. One of the exciting moments of this conference was Duifhuis's discussion of the motor-active outer hair cells, an input management system which may, in part, account for the non-linearities which Goldstein so beautifully modelled. In short, some management of input is already built into the modular parts of our senses.

But non-modular feedback will, in many cases, be a main contributor to accurate convergence. Due to the perceiver's attentional control, a selection of potentially relevant input is made. Wright, in her paper, went so far as

saying that there are explicit strategies of non-selection or evasion, which by default optimize the relevant input. She suggested that this is rather unexplored theoretical territory in cognitive psychology. But is it really? As Bouwhuis pointed out, existing models of goal/action hierarchies are relevant here. Rosenbloom's work shows that this is by no means virgin area in psychology.

One would like to know better to what extent the management of input affects or is even essential for accurate convergence. Much work in visual perception is based on the tacit assumption that input management is irrelevant. Leeuwenberg's theory, for instance, is ignorant of such factors. And indeed, it would be very pleasant if major areas of perceptual theory could be saved from input management factors. But even for these classical areas of perceptual theory there are inconvenient phenomena. Anstis, for instance, could show that the direction of apparent motion can depend on where the eye fixates. And Juola reminded us of the important fact that no movement is seen when the eye is actively moved. No way to account for such phenomena by a structural information theory. The perceiver's management of input is crucial here. And it should be added that input management is probably a doubly important factor in the acquisition phase of a perceptual skill. How does the learner scan the input? Is there a systematic interaction between the developmental state of the perceptual system and the input it selects? These are no small questions.

But the problems become even more intractable when the input management is not in the hands of the perceiver. This is often the case in speech perception. The listener usually has little control over what he wants to listen to. It is, rather, prepared for him, and presented in some pre-planned order. It is the speaker who is supposed to optimize the input. And this optimization affects different levels of the input. Lindblom's 'adaptive variability' is such an input management strategy. It will, for instance, affect the size of the vowel space the listener becomes confronted with, as well as the amounts of assimilation and coarticulation he will have to deal with. In response, the addressee can exert a certain amount of control by saying 'What?', or by asking for more specific repair. Real life speech perception to some extent relies on these corrective possibilities. Will a model of speech perception ever be complete if these kinds of input management are not taken into account? I doubt it.

Similar remarks can be made with respect to text comprehension in general. Bunt's dialogue system has been the most explicit working model of mutual input management at this symposium. Given his goal, an inter-

locutor constructs the next input text on the basis of a running account of the state of discourse and the state of dialogue. He systematically directs the addressee, usually in a sequence of turns, to come up with the correct interpretation of his intention.

In short, there are large areas of perceptual modelling where input management is or should be a crucial component of the theory if it aims at ecological validity.

6 Issue 5: What is general and what is specific in perceptual architecture?

The last concern I would ask you to share with me, is the ever-existing tension between general and specific explanation in perception. They are styles of theorizing which have been with us since Plato and Aristotle and from which we will not disentangle ourselves for a long time to come. The approach of general explanation involves the statement of general principles of perceptual organization, and the assumption of a fairly general processing architecture underlying the different perceptual modalities. At this conference we have seen two typical instances of this approach. Leeuwenberg's structural information theory is to a large extent modality-independent. It invokes general principles of perceptual organization, in the same vein as Gestalt Psychology used to do. By implication, somewhat less in the forefront of the work, it assumes the existence of general processing algorithms which will parse perceptual input according to these organizational principles. The other case is Rosenbloom's work. Its explicit goal is to create a general theory of intelligent behaviour. The basic architecture of the system is quite independent of the specific tasks it performs. It is a *uniform* architecture. What there is in terms of specificity is acquired through chunking, but the principles of chunking are, again, of an entirely general nature. In both these cases we have to do with a purposeful generalistic approach. There is a genuine belief in the existence of powerful organizational principles of a general kind.

Specific explanation is often done by default, not because of a principled view on perceptual organization. If one works on a model of the cochlea, as Goldstein and Duifhuis do, one is by necessity working on a specific model. There are no hair cells in the retina. In Anstis's work on movement and luminance detectors, he is equally creating a specific model. All this is obvious and unobjectionable.

But the search for specific explanation can be as principled as the search for general explanation. Proponents of modularity claim that input systems are specific. They are specific in the kind of input they accept, they are specific in their principles of organization, they are specific in the kind of output representations they compute, and they are specific in their patterns of break-down. In addition, they are supposed to be implemented in specific neural tissue. The architecture of the auditory cortex is entirely different from the organization of the visual cortex, according to this view.

Speech perception has an interesting place in this modular picture. It comes through the ear, but it is not just auditory perception. It has its own principles of organization. There is something like a 'speech mode' (Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967) which treats acoustic input 'as if it were speech', i.e. imposing a set of speech-specific perceptual categories through which the signal can be mapped onto a phonological code, for instance a word form. A possible parallel case in visual perception could be a 'face mode', a specific way of dealing with the categorization of faces and of facial expressions. For both the speech mode and the face mode the reasonable conjecture has been made that they are subserved by specific neural tissues.

The latter two examples should guard us from making the oversimplification that input systems are obviously specific, because the sense organs are specific. The speech mode and the face mode are post-cochlear and post-retinal, respectively. Modularity is, therefore, by no means a trivial conjecture.

The tension between general and specific approaches to perception is, in my view, a healthy and stimulating one. But it would be too optimistic to predict that the dialectics will once resolve itself in a grand synthesis. That kind of Utopia is only given to political systems.

7 Epilogue

This completes my five general issues in perception. I began by making a distinction between homogeneous and heterogeneous scientific meetings. We are completing a heterogeneous one, and we should be proud of having survived the tensions this created. But let us not forget that the Institute for Perception Research IPO has already managed to survive no less than 30 years of such tension. It is apparently possible for physicists, engineers, linguists, psychologists, phoneticians, logicians and computer scientists to

cooperate peacefully in the study of human perception. IPO has set an example hardly matched anywhere in the world. It should be congratulated. And I am sure I am expressing the feelings of all of us when I thank the Institute, Herman Bouma and his magnificent dedicated team for having organized this exciting symposium.

References

Groot, A.D. de (1946) *Het denken van den schaker.* Amsterdam: Noord-Hollandse Uitgevers Maatschappij (In translation: *Thought and choice in chess.* Den Haag: Mouton, 1965).

Levelt, W.J.M. and Flores d'Arcais, G.B. (1987) Snelheid en uniciteit bij lexicale toegang. In: H.F.M. Crombach, L.J.Th. van der Kamp and C.A.J. Vlek (eds), *Ontwikkelingen rond model, metriek en methode in de gedragswetenschappen*, Lisse: Swets & Zeitlinger.

Levelt, W.J.M. and Schriefers, H. (1987) Stages of lexical access. In: G. Kempen (ed.), *Natural language generation. New Results in artificial intelligence, psychology and linguistics*, Dordrecht: Martinus Nijhoff.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. and Studdert-Kennedy, M. (1967) Perception of the speech code, *Psychological Review, 74*, 431-461.