AUDIO-VISUAL SENSORY INTERACTIONS AND THE STATISTICAL COVARIANCE OF THE NATURAL ENVIRONMENT*

(Invited paper)

PACS REFERENCE: 43.66.Ba

Zetzsche, Christoph; Röhrbein, Florian; Hofbauer, Markus; Schill, Kerstin
Institut für Medizische Psychologie, Ludwig-Maximilians-Universität München
Goethestr. 31
80336 München
Germany
Tel: +49 89 5996 218
Fax: +49 89 5996 615
E-mail: chris@imp.med.uni-muenchen.de

ABSTRACT

We show that a mobile observer in a natural environment receives systematically co-varying signals in his different sensory modalities. An independent, modality-specific processing - as assumed in classical theories of perception - would hence be sub-optimal. Rather, information theory predicts that the system should use a statistically optimised joint processing strategy. We tested this by measuring the two-dimensional just-noticeable difference (jnd) curves for basic visual-auditory stimulus configurations (a patch of light combined with a 1kHz tone). The forced-choice task was to detect any change in this configuration, irrespective of modality. The resulting two-dimensional jnd-curve cannot be explained by an independent, modality-specific processing. In particular, the sensitivity increase for the "ecologically relevant" joint auditory-visual increments or decrements is much higher than the usual probability summation effects. This points to a direct neural integration of visual and auditory information at an early stage.

INTRODUCTION

The properties of perception are usually studied for only one modality at a time, and in strict isolation from the remaining modalities. The assumption behind this research strategy is that the processing of sensory signals is performed in independent, modality-specific channels. However, it is questionable whether the signals received by the different modalities are completely independent of each other. But if they are not independent, would it then really be a good strategy to keep the processing channels for these signals strictly separate?

Surprisingly, it are recent advances in the understanding of *unimodal* sensory processing, which cast the strongest doubts on the appropriateness of the “modality independence hypothesis”. According to these results, the major driving force for the specification and development of the sensory neural computations is just the statistical dependence between individual sensory messages. The key for understanding this relationship is given by information theory: if sensory messages exhibit statistical covariations, then it is suboptimal to process them independent of each other and it is more efficient to transform them into a new code which is derived from appropriate *combinations* of these correlated signals. A prominent example for such a strategy are the receptive fields of neurons in the visual system. These receptive fields represent just the optimal combinations of the signals of the individual photoreceptors for an efficient exploitation of the statistical properties of the visual environment (for review see, e.g., [1]).

Now if such an information-theoretic optimisation principle would be a *universal* principle of sensory information processing (and not only valid for intramodal processing) there would immediately result a specific prediction: if there exist systematic statistical covariations between

* Supported by DFG, SFB 462 Sensomotorik, B5 and GRK 267 Sensorische Interaktion

the signals in the different modalities, then their processing should *not* be independent. Rather, there should exist specific *combinations* - i.e. cross-modal interactions - and these interactions should be found to be matched to the statistical multimodal covariance structure as caused by the typical environmental and behavioural properties. This prediction is tested in this paper.

The first point to check is obviously, whether there exist such systematic statistical relations between the modalities. A direct measurement would be no trivial task, but we can find a first approximation by means of a computer simulation. For this we compute intermodal statistical co-variations between visual and auditory signals, which result from the natural scenario of an observer who moves around within an environment of other moving objects (Fig 1). Each object has a characteristic intrinsic size and sound level, and the relative movement between objects and observer is variable (observer stands and objects move, both move, only observer moves). The simulation is run in discrete small time-steps for various random configurations and the perceived size and loudness of the objects at the observers' position is registered for each point in time and plotted in a two-dimensional auditory-visual coordinate system (Fig 1, right).

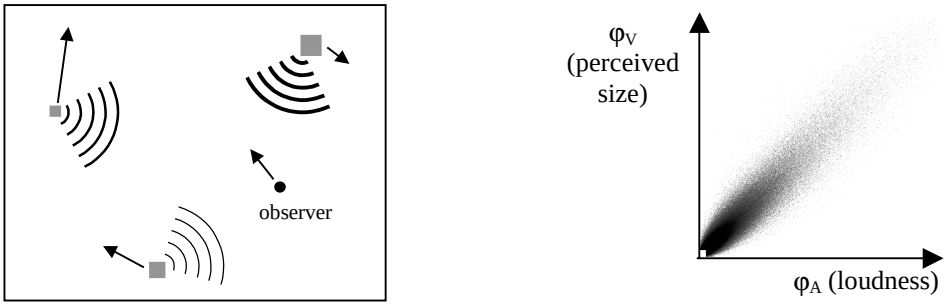


Fig 1: Simulation of the statistical visual–auditory covariances of an observer moving around in an environment with other moving objects. Left: Schematic setting. Right: Resulting statistical distribution of the perceived size and loudness values. Points in the upper right corner represent objects, which are perceived as big and loud, whereas points in the lower left corner correspond to tiny and mute objects.

We found a clear statistical correlation between the perceived size and loudness values which is largely independent of the special parameters used for the statistical distributions of the movements and of the intrinsic object size and sound-level values. A similar statistical covariation can hence be expected to result for an actual human observer in a real environment.

EXPERIMENTS

Given that the visual and auditory signals are actually not independent, the next step is to design an appropriate experiment in order to test for the predicted intermodal interactions that are matched to the structure of these statistical covariances. Note that experiments on multisensory interactions are by no means a novel issue (though they represent still a marginal, if growing fraction of sensory research). However, although there exist a number of experimental findings which suggest certain types of intermodal interactions (for review see, e.g. [2]-[5]) the fundamental logical status of such interactions, as well as their locus, and their computational structure, are still topics of controversial debate (e.g. [5]-[12]). The most basic conflict in the interpretation of the experimental data is whether these behaviourally observed interactions have an “early” or a “late” basis. “Hardliners” from the more traditional position would insist that it is still correct to assume that the sensory information is processed essentially separate within each modality, because all the observed interaction effects can be attributed to either (i) interactions on some higher, more abstract and post-modal stage, (ii) a mere interpretation or decision bias, as opposed to a true early interaction, (iii) to forewarning effects or (iv) to mere statistical advantages. A typical example for a late, high-level effect is the improved identifiability of the gender of a person if both visual and auditory cues are available. A typical example for a decision bias can be seen in the ventriloquist effect [6] (but see [10]). A forewarning effect [13] can be a problem in the unique interpretation of reaction-time advantages for bimodal stimuli (the redundant target effect) since the observed differences are often quite small. An interesting variant of the late interaction hypothesis for reaction time measurements is the attribution of the observed interaction effects to the motoric processing stages [8]. And finally, "statistical facilitation" can cause additional complications since the mere availability of multiple signals at the decision stage can lead to a statistical advantage for the correct reaction, as assumed in probability summation or race models [9].

Against this sceptic position, several authors have proposed arguments and experiments in order to prove that “true”, or “early” intermodal interactions (also known as "neural integration" or "coactivation”) do actually exist, i.e., cannot be reduced to any of the late effects, e.g., [7], [9],[10]. However, since the reasoning is often quite complicated it has not yet succeeded in bringing the controversial discussion to a final and definitive end. It would hence be desirable to find simple and straight-forward demonstrations of early inter-sensory integration effects, which can substantially challenge any version of the conservative late-interaction hypothesis. The following experiment is one attempt towards the provision of such a demonstration.

What would be desirable properties of such a demonstration? Ideally, the experimental design should already exclude all the mentioned “non-genuine” interaction effects, i.e. high-level and late interactions, bias, fore-warning and probability summation. Reaction time experiments are less well suited, since it is difficult to exclude fore-warning effects, motor-level integration and statistical facilitation. In spatial position judgements, like the ventriloquist effect, it is difficult to exclude a late bias. This leaves accuracy or performance measures in a signal detection sense, which are typically aimed at the identification of the processing properties of early processing stages. If we use a two-alternative forced choice design (2AFC), we can exclude any bias or related late interaction effects right from the start. This leaves probability summation as the only remaining source for a possible confounding of early and late interaction effects. The solution of this last problem can profit from the information-theoretic perspective, since the predicted preference of *specific* signal combinations, which are matched to the statistical covariance structure implies that the cross-modal interaction effects should have no *uniform* structure (i.e., there should be no “blind” or unspecific pooling of information from different sensory channels). This gives us a chance for finding - within one and the same experimental setting - such cross-modal signal combinations that lead to true interactions, and also other combinations, which lead only to probability summation. It is then not necessary to proof that the observed interactions really exceed a difficult to define theoretical baseline, but it is logically sufficient if the different interaction effects that occur within the experiment are clearly distinguishably, since they can then obviously be not all reduced to the same late statistical facilitation effect.

Methods

Based on the above considerations we designed a suitable visual-auditory discrimination task. If the neural computations are really adapted to the statistical covariance structure shown in Fig 1 then an efficient encoding can be obtained if the system provides specialized neural resources for the representation of those bimodal signal combinations which allow for a precise distinction of the statistically most typical visual-auditory changes, i.e. for the correlated bimodal changes along the main diagonal. No specialized resources are required, on the other hand, to evaluate the statistically untypical changes, i.e. those bimodal changes in which one component increases while the other one decreases.

To test this prediction we independently varied the size and the loudness of a bimodal visual-auditory stimulus. We measured the two-dimensional bimodal jnd-curves for the detection of joint bimodal and unimodal changes $\Delta\varphi_A$; $\Delta\varphi_V$ of the loudness and/or size of a reference compound stimulus (Fig 2). This enabled us to test various possible combinations $\Delta\varphi_A$; $\Delta\varphi_V$ including (i) ecologically typical cases in which the subjects can have the impression of a sound-emitting object, which can come closer or can move away in space, (ii) less typical cases where one component changes strongly while the other one changes only a little bit or stays constant, and (iii) untypical cases in which one component increases while the other one decreases.

We used a temporal 2AFC design in which we presented two subsequent size/loudness combinations in a first interval, and after a short pause another two combinations in a second interval. The subjects’ task was to decide in which of the two intervals the stimulus configuration has changed, irrespective of whether this change was due to a change of the visual signal, or of the auditory signal, or of both. For this they had to give an unspeeded response by pressing one of two buttons.

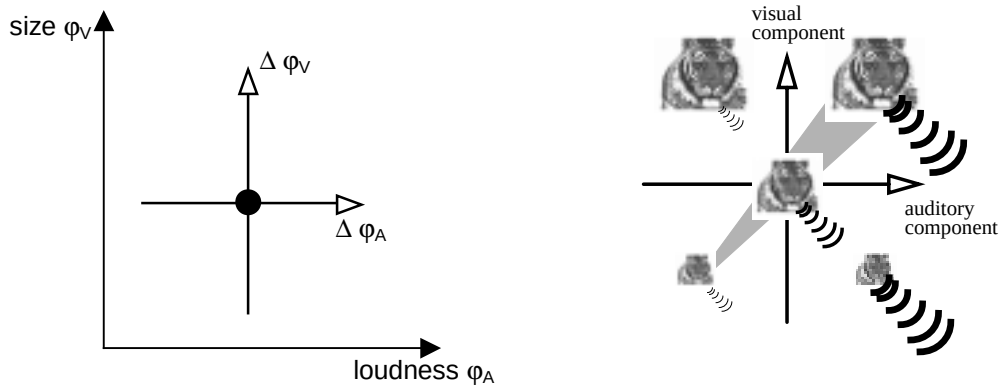


Fig 2: Two-dimensional visual-auditory discrimination paradigm. The left side shows the respective axes, with the reference compound stimulus denoted by the thick circle. The right side shows a schematic illustration of the meaning of the possible two-dimensional changes.

The visual stimulus was a grey square (2.5 deg vis) on black background, displayed on a monitor in a semi-darkened room. The auditory component was a 1 kHz tone and was presented via headphones. Both components were presented simultaneously for 400 msec with an ISI (inter stimulus interval) of 200 msec. The thresholds were determined with an adaptive procedure, which lead to about 40 presentations per threshold. We determined 28 combined auditory-visual difference thresholds for 14 subjects, and measured two thresholds within one trial in carefully selected combinations in order to ascertain that the subjects attend to both signal components all the time. We performed control experiments with ISI's of 400 and 800msec, and with asynchronous auditory and visual components (SOA 800 msec).

Results

The results are plotted in Fig 3. The figure shows clearly that the bimodal sensitivity distribution exhibits the predicted match to the statistical covariance of Fig 1. As can be seen from the lightly shadowed regions the obtained gain for correlated component changes $\Delta\varphi_A$; $\Delta\varphi_V$, i.e. for the ecologically typical signals with simultaneous increments or decrements for both size and loudness, is substantially higher than the gain in the statistically non-typical quadrants, where an increment in one modality is combined with a decrement in the other modality.

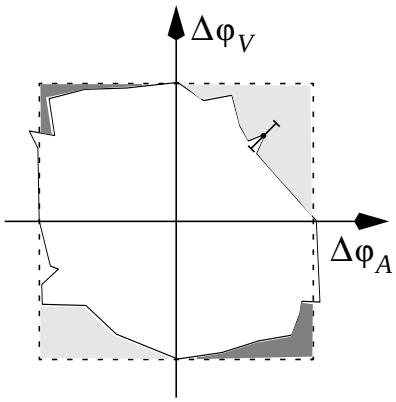


Fig 3: Two-dimensional visual-auditory jnd-curves. Points on the jnd-curve represent measured relative thresholds averaged over 14 subjects. The data were scaled to the four unimodal thresholds on the abscissa and ordinate, since we are interested in the relative performance. The oblique error bar plotted for one of the joint auditory-visual increment combinations denotes the average standard deviation.

As argued before, the mere difference between the quadrants is already logically sufficient to exclude a complete reduction to a late-stage probability summation effect, and is therefore a proof for the existence of true intermodal interactions on an early processing level. However, the sensitivity increase in the "ecological" quadrants is so strong, that it can also be distinguished from probability summation effects in absolute terms, as will be shown in the following section.

With the values tested we found neither an influence of the presentation mode (simultaneous or sequential) nor did the extension of the ISI influence accuracy.

MODELS

Two basic models of multisensory interaction can be distinguished in the given bias-free signal detection context. Both schemes presume the same sort of input stage with separate, modality-specific signals but differ in the further processing of these signals, i.e. in the way, how the multisensory information is pooled. As initial stage for both models, the two stimuli S_1 and S_2 lead to different activation on some psychological coordinates (Ψ_A ; Ψ_V) in the auditory and the visual channel. In the "separate activation model" (Fig 4, upper part) an independent decision is then made separately in each channel. The observed response of the subject results from an OR-combination of these two modality-specific decisions in the final decision stage, and the joint probability of discriminating the compounds becomes hence a product of the probabilities for reaching the unimodal thresholds. The overall probability for a correct decision is hence slightly raised (or the stimulus changes required for a fixed percentage of correct responses are reduced below the unimodal values) if multisensory stimulation is applied. Since an observed gain is here attributed to a pooling of probabilities, this scheme is called probability summation model (or race model, if reaction times are considered). The jnd-curve predicted by this model would be roughly rectangular with rounded corners.

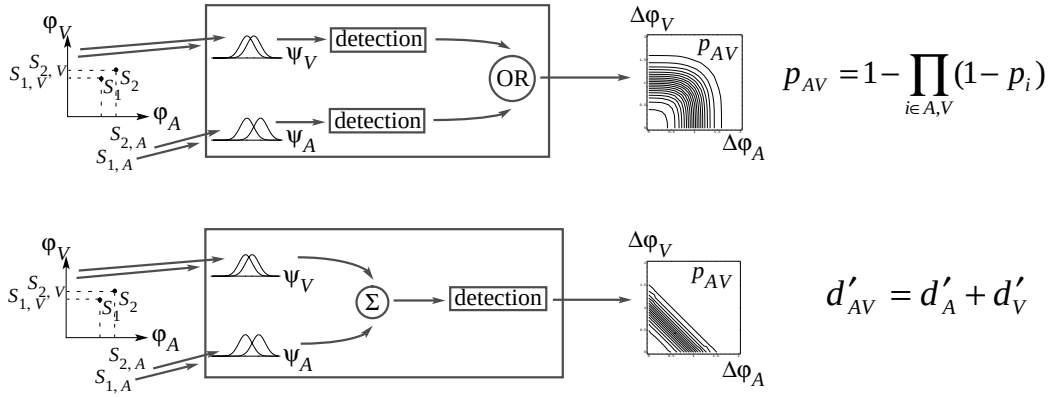


Fig 4: Two basic models of multisensory interaction: Upper figure: Late interaction in the separate activation or probability summation model. Lower figure: Early cross-modal interactions in the neural summation (coactivation) model.

A quite different shape of the jnd-curve results from the "neural summation model" (or "coactivation model"), as shown in the lower part of Fig 4. Here it is assumed that the information from different modalities actually converges on an early stage prior to the decision. Instead of mere pooling of probabilities a direct physiological-neural summation takes place, and based on the resulting neural output value one final decision is obtained. This sort of pooling amounts to a linear combination of the input variables, and the resulting two-dimensional threshold curve would therefore have a rhomboid shape.

Interestingly, our experiment shows both types of interaction effects. The sensitivity gains we observed for the ecologically typical stimulus combinations (first and third quadrant) are in good agreement with direct neural summation, whereas the gains in the other quadrants come closer to an independent processing with mere statistical pooling. Thus neither a probability summation model nor a neural summation model alone can account for the structure of the data. We hence propose a model, which allows for a selective interaction of the auditory and visual channel by postulating two sorts of processing streams, an ON and an OFF stream (like those found in the visual system [14]. Signals with equal signs are linearly combined and enable thus decisions with a significantly increased sensitivity. If the signs differ, however, independent decisions are computed for the two modalities based on the respective ON and OFF-subsystem, and there is only a slight facilitation effect due to statistical pooling.

DISCUSSION

Our experimental results provide clear evidence for a genuine integration of multi- sensory information at a low to intermediate level. Our analysis was based upon an information-theoretic approach which interprets sensory processing as an optimal adaptation to the multisensory statistical covariance structure that is generated by the typical behavior in a natural environment. In a simulation of such natural conditions we found substantial covariations of auditory and visual signals. In order to test the prediction that there should exist corresponding

bimodal interaction effects, we designed a novel 2AFC experiment for the measurement of two-dimensional visual-auditory jnd curves. This design excludes the possibility of bias and other late interaction effects. As predicted by the hypothesis, we found an *early, highly selective multisensory integration* for the "ecologically relevant" stimulus combinations. Our results are therefore a clear counter argument to the classical idea of basically separate, modality-specific sensory channels, which can only interact in form of bias effects or by a *late, unspecific statistical pooling* at a higher-level decision stage.

The evidence for the localization of this integration process within the neural architecture of the brain is as yet inconclusive. The colliculus superior is definitely a site of multimodal integration [3], but its relevance for perception is unclear. At the cortical level, certain regions link multisensory information for perception and language [15][16] like area LIP in posterior parietal cortex, AES in cat lateral temporal cortex and STS, which receive input especially from sites known to have visual looming detectors. However, it remains to be shown that these sites can perform a true multimodal signal integration, rather than only a gating function or attentional control. Recent fMRI studies indicate that no special neural site for the explicit summation of signals from the two modalities might exist at all. Rather, there could only exist a reciprocal amplification between the unimodal cortices [5]. This may also be consistent with various reports about the excitability of neurons in the primary sensory cortices by stimuli from other modalities, e.g. [17]. The final identification of the neural substrate of the observed intersensory interactions has thus to await more detailed neurophysiological studies.

REFERENCES

[1] Zetzsche, C. ; Krieger, G. (2001), Nonlinear mechanisms and higher-order statistics in biological vision and electronic image processing: review and perspectives. J. of Electronic Imaging 10(1), 56-99.

[2] Welch, R. B; Warren, D.H. (1986) Intersensory interactions. In: Handbook of Perception and Human Performance, Vol.1: Sensory Processes and Perception, K. R. Boff and L. Kaufmann and J. P. Thomas (Eds.), Wiley, NY, Ch. 25, 1-36.

[3] Stein, B. E. and Meredith, M. A. (1993). The merging of the senses. MIT Press, Cambridge, MA.

[4] King, A.J; Cavert, G.A. (2001) Multisensory integration: perceptual grouping by eye and ear. Current Biology 11, R322-R325

[5] Driver, J.; Spence, C. (2000) Multisensory perception: beyond modularity and convergence. Current Biology 10, R731-R735.

[6] Choe, C.S.; Welch, R.B.; R.M. Gilford, R.M.; J.F. Juola, J.F (1975) The ventriloquist effect: visual dominance or response bias, Perception & Psychophysics 18, 55-60.

[7] Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. /. Exp. Psycho/. Hum. Percept. Perform. 17(1), 160-169.

[8] Giray, M.; Ulrich, R. (1993). Motor coactivation revealed by response force in divided and focused attention. J. Exp. Psychol. Hum. Percept. Perform., 19(6): 1278-91.

[9] Townsend, J.T. ; Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. J. Math. Psychol., 39(4):321-359.

[10] Bertelson, P.; Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. Psychonomic Bulletin & Review 5, 482-489.

[11] Calvert, G.A.; Brammer, M.J.; Iversen, S.D. (1998). Crossmodal identification. Trends Cognit. Sci. 2:247-253.

[12] Meyer, G.F. ; Wuerger, S.M (2001). Crossmodal integration of auditory and visual motion signals .NeuroReport, 12(11), 2557-2560.

[13] Nickerson, S. (1973) Intersensory facilitation of reaction time: energy summation or preparation enhancement? Psych. Rev. 80, 489-509.

[14] Schiller, P.H. (1992) The ON and OFF channels of the visual system. Trends in Neurosciences 15(3), 86-92.

[15] Wallace, M.T.; Meredith, M.A.; Stein, B.E. (1992). Integration of multiple sensory modalities in cat cortex. Experimental Brain Research 91(3), 484-488.

[16] Downar, J., Crawley, A. P., Mikulis, D. J., and Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. Nature Neuroscience, 3(3):277-283.

[17] Spinelli, D.N., Starr, A., and Barrett, T.W. (1968). Auditory specificity in unit recordings from cat's visual cortex. Experimental Neurology 22, 75-84.

[18] Ref2 sd sd sd sdg dg sd sdg sdg sdg sdg sdg dg sdfg sd dfggsdsd gsd fgsdggsgsd gd g ds sdf sdfg sdfg sdg sdg sdg sdg sdg sdg sdg dg

[19] Andersen, R. A. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. Phil. Trans. R. Soc. Land. B, 3.52:1421-1428.

[20] Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., and David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. NeuroReport, lO(12):2619-2623.

[21] Colonius, H. and Townsend, J. T. (1997). Activation-state representation of models for the redundant-signals-effect.

[22] In Marley, A. A. and Mahwah, N. J., editors, Choice, decision and measurement, pages 245-254.Lawrence Erlbaum.

[23] Craig, A., Colquhoun, W. P., and Corcoran, D. W. J. (1976). Combining evidence presented simultanously to the eye and the ear: A comparison of predictive models. Perception and Psychophysics, 19(6):473-484.

[24] Fishman, M. C. and Michael, C. R. (1973). Integration of auditory information in the cat's visual cortex. Vision Research, 13:1415-1419.

[25] Giard, M. H. and Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. Journal of Cognitive Neuroscience, 11(5):473-490.

[26] Graham, N. (1989). Visurd Pattern Analyzers. Oxford University Press, New York.

[27] Green, D. M. and Swets, J. A. (1966). Signal Detection Theory and Psychophysics. Wiley.

[28] Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., and Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. J. Exp. Psychol. Hum. Percept. Perform., 20(1):131-153.

[29] Loveless, N. E., Brebner, J., and Hamilton, P. (1970). Bisensory presentation of information. Psychological Bulletin. 73(3):161-199.

[30] Marks, L. E. (1978). The unity ofthe senses. Academic Press series in cognition and perception. Acadamic Press.

[31] Mordkoff, J. T. and Yantis, S. (1991). An interactive race model of divided attention. J. Exp. Psychol. Hum. Percept. Perform., 17(2):520-538.

[32] Mulligan, R. M. and Shaw, M. L. (1980). Multimodal signal detection: Independent decisions vs. integration. Perception and Psychophysics, 28(5):471-478.

[33] Raab, D. (1962). Statistical facilitation of simple reaction time. Transact. N. Y Acad. qfSci., 43574-590.

[34] Sakata, H., Taira, M., Kusunoki, M., Murata, A., and Tanaka, Y. (1997). The parietal association cortex in depth perception and visual control of hand action. Trends in Neuroscience, 20(8):350-357.

[35]

[36] Sumby, W. H. and Polack, I. (1954). visual contribution to speech intelligibility in noise. J. Acoust. Sot. Am., 26:212-215.

[37] Treutwein, B. (1997). Yaap: Yet another adaptive procedure. Spatial Vision, 11: 129-134.

[38]

[39] Watanabe, J. and Iwai, E. (1991). neuronal activity in visual, auditory and polysensory areas in the monkey temporal cortex during visual fixation task. Brain Research Bulletin, 26:583-592.

[40] Zetzsche, C. and Krieger, G. (1999). Nonlinear neurons and higher-order statistics: new approaches to human vision and electronic image processing. In Rogowitz, B. and Pappas, T., editors, HLIIJ~I vision and Electronic lmnge Processing, volume 3644 of Proc. SPIE, pages 2-33, Bellingham, WA.

[41] Zwicker, E. and Fastl, H. (1999). Psychoacoustics. Springer, 2 edition.

========== RESTE INTRO ==========
==========

The main argument in favour of this "perceptual independence hypothesis" is that evolution has provided us with a set of specialized receptor mechanisms to obtain information about our environment. The origin of this specialized subsystems may be sought in the need for a detection of specific events, for which information is only emitted in restricted subbands of the spectrum of physical effects, or it may have its cause in the mere impossibility of designing a "universal", i.e. ultra-broadband receptor mechanism which covers the whole range of environmental information sources.

However, while it is certainly true that there exist many situations in everyday life in which we have to rely on one specific modality, it is also evident that there exist at least as many situations in which a single object or event leads to systematic covariations in several sensory channels. These effects, together with more philosophical arguments for a multimodal object representations as basis of our unitary experience (e.g. Marks78), have recently lead to an increasing interest in inter-modal sensory interactions (e.g., cite review??). However, empirical studies of multisensory processes represent still a marginal, though growing fraction of the entire area of sensory research, and the fundamental logical status of the nature of multisensory interactions is still a highly controversial issue.

The core of the problem is the as yet unidentified locus and computational structure of multisensory perceptual interactions (CaBrIv98; DoCrMiDa00; Townsend and Nozawa, 1995).

======

(also known as "neural integration" or "coactivation", e.g. miller91 GreSwe66, LoBrHa70)

=======

Evidence that multisensory interactions cannot always be reduced to a statistical combination of the outputs of independent channels has also been found in other investigations. For example, reaction time reductions for combined stimuli tend to be stronger than the predictions of an independent race model (Mi91, Hughes94). However, this issue is still under debate. The inconsistent reaction time differences are usually quite small, and the proof of inconsistency with an independent processing model is not trivial (e.g. ToNo95, CoTo97, GiPe99).There are also claims that the observed interactions may take place on a motoric rather than a perceptual level (e.g., Giray, Ulrich, 1993 ??neuere, bessere ref??). This could be especially relevant for saccadic reactions, which may be based on a separate, low- level sensorimotor path mediated by the superior colliculus (StMe93). Intermodal interactions have also been observed on higher cognitive stages, for example in the improvement of speech intelligibilty by combined audio-visual cues (cite??), but these high-level effects can not be of help for a clarification of the issueof multisensory integration on an early perceptual level.<<

======

Common sense suggests that such multimodal statistical covariations may easily occur, for example the single event of the appearance of a new object may easily generate certain signal covariations in different modalities, and there is already an increasing number of reports which indicate some sort of intermodal interaction (for review see, e.g. ??).

==========

==========

Ideally, the task should contain, within one setting, conditions which lead to true sensory integration effects and conditions which cause no integration but only probability summation effects
Not an unconditional pooling of information from two modalities but selective integration depending on conditions
Clearly distinguishable from probability summation effects
Exlude decision bias and other late cognitive factors
The task should have a forced choice design in which the which is orthogonal to includes any type of bias right.

==========

It must be mentioned that the few studies available so far (e.g. CrCoCo76, MuSh80), like the old

vigilance studies reviewed in LoBrHa70, have not succeeded in yielding clear interaction effects. However, this should not prohibit us in searching for some new multi-modal performance task which reveals the desired effects.

=========

With respect to finding a suitable task we can consider a basic argument in favour of interaction effects somewhat closer. A critical argument against a strictly separate processing is that the signals in the different channels are not always independent of each other. It is true that there exist extreme cases where we hear something without seeing anything (at night, for example), and vice versa, and that in certain cases what we see has no relation whatsoever to what we hear (e.g. if an object is occluded by another one). However, in many cases there is some relation between what we hear and what we see, such that there exists at least a *statistical dependence* between the signals in the two sensory channels. Such statistical covariations have recently proven crucial for the understanding of sensory processing within one modality. In particular, the peculiar neural interactions between the signals from the individual receptors in the eye, which are realized by the retinal and cortical neurons have shown to be efficient means for the exploitation of the statistical redundancies of the natural environment (for review see, e.g., ZeKr99). If an exploitation of the statistical covariances in sensory messages is a basic principle of neural processing, such an exploitation of the covariances between the signals from different modalities requires inevitably the provision of specific multisensory signal combinations, i.e. a “true”, or early interaction in the sense of the above discussion. Our study thus aims at the provision of unambiguos evidence for such a "true" integration of multimodal information at a low to intermediate processing level, which expoits the ecologically relevant covariance structure of multisensory signals.

==========

==========

This characteristic natural covariation structure gives us the basis for the design of a suitable experiment for the demonstration of early intermodal interactions. If the assumption that the neural computations are adapted to this statistical covariance structure is correct, then we have to expect that the system provides specialized neural resources for the encoding of those bimodal signal combinations which allow for a precise distinction of the correlated visual-auditory changes (i.e. changes along the main diagonal), while it does not provide specialized neural resources to evaluate the statistically untypical bimodal changes in which one component increases while the other one decreases.

==========

Essentail question: what is combined, and how is it ombined. It is clear that here a strong ecological component can be expected to come into play ... However, as far basically restricted to spatial and temporal coincidence.

but it is obvious that these sensory mechanisms are now used in most cases in an an integrated rather than a separated fashion. Although multisensory convergence and integration are

bias vs. improvement (preparation enhancement or alerting effect, central summation of energies)

-
- ventriloquism
- evoked potentials
- synesthesia
- compensatory eye movements due to visual and vestibular inputs
- Piaget and Helmholtz: senses are separated at birth and become interconnected through experience

- McGurk

Multisensory integration can be expected to play a significant role in nearly any kind of behavior.

- either neural, than mostly low level, e.g. colliculus, or behavioral, than mostly high-level, e.g. speech understanding

reflexive reactions
typical definition of coherent stimuli: spatially and temporally coincident.
hunting as well as predator avoidance

- RT very small effect
- separate (independent) activation vs. coativation (integration, neural summation (Mill91))

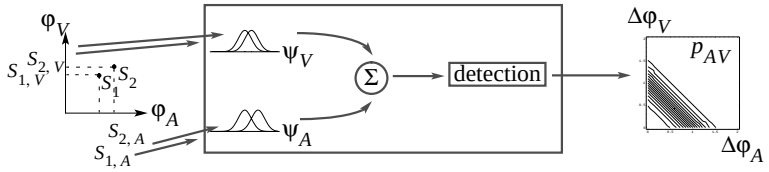
1. objects in a natural environment are often perceivable through several modalities
2. movement of the object or of the observer causes systematic covariations in the input channels
3. can this redundant information be utilized? more general: how is information from different modalities integrated? (bekannt: multisensory integration can improve speach perception and localisation)
Literatur 'intersensory interaction' with task-relevant redundant info.:
1. facilitation effects from vigilance studies / sonar performance (=event detection), but no systematic investigation
2. intersensory bias: conflicting information (e.g. McGurk). typical result: visual dominance (e.g. ventriloquist).
3. reaction time experiments: "redundant target effect".
4. very view studies on detection / discrimination accuracy

========== RESTE EXPERIMENTS ==========
>>Überhaupt erwähnen??: In a further experiment we measured the two-dimensional threshold curve for bisensory intensity differences. Here the bimodal stimuls compound consisted of a bright lightspot of 3.5 min arc and of a 1kHz tone. Presentation time was 700 msec with 900 msec (ISI). We determined 16 combined auditory-visual difference thresholds.<<

=========

The results are plotted in Fig 3 as difference threshold curves around the reference compound (which is used as origin of the coordinate system). Positive values indicate increments and negative values decrements.

==========



$$d'_{AV} = d'_A + d'_V$$

Fig 5: Early cross-modal interactions in the neural summation (coactivation) model.

==========
The data show clearly that the subjects show substantial intersensory facilitation effects. As can be seen from the lightly shadowed regions the obtained gain for correlated component changes $\Delta\varphi_A$; $\Delta\varphi_V$, i.e. for the ecologically typical signals with simultaneous increments or decrements for both size and loudness, cannot be explained as a mere statistical effect in the sense of probability summation. Rather it comes close to a true integration of the two signal components, i.e. to bisensory summation of the incoming signals at a relatively early stage in the system. This is further supported by the fact that the amount of facilitation is clearly dependent on the quadrants. The non-typical quadrants with a combination of an increment in one modality with a decrement in the other one show only a minor increase in sensitivity, which is compatible with the assumption of an independent modality-specific processing and a late probability

summation effect.

==========

======== RESTE DISCUSSION ==
Other evidence for neural summation (hier?):
- response force (Giray, Ulrich, 1993)
- saccadic reaction times (Nozawa et al., 1994)
- neuro: CS (..Baddeley..)

evidence for interactive coactivation:
- RT: pitch + position (Miller, 1991)
- event-related brain potentials (Schroeger, Widmann, 1998)

neurophysio data / convergence areas:
- colliculus superior (primate: WaWiSt 1996)
- cortex: multimodal sensory association areas, hier v.a. posterior association area (=pario- temporal, e.g. STS)
- loomimg detec. sound mov. areas

A: neural substrate for sensory integration: cortex

1. early stages / primary sensory cortices:
1.1 reciprocal amplification in unimodal cortices (ref. Calvert)
1.2 “auditory neurons” in visual cortex area 17, 18 (diverse refs)

2. higher-order cortices:
interaction bei cells in TEO (visual area) and in AA (auditory), ref. Watanabe.

3. association cortices:
2.1 grosse Bereiche des Cortex gelten als multimodal association areas; dazu passt auch: multimodal distributed cortical network (ref. Downar et al.)
2.2 fuer uns bes. relevant (und auch schon riesig): posterior association area (pariotemporal lobe) which links information from several sensory modalities for perception and language. Dazu gehoert area LIP im posterior parietal cortex (v. a. stimulus location, e.g. ref. Anderson), STS (v.a. speach, input auch von MT) und AES im lateral temporal cortex (ref. WallaceStein).

B. motion processing:

1. visual: loomimg detectors in primate MST: respond to change in image size + change in disparity (ref. e.g. Sakata 1997, hab ich einiges)
2. acoustic motion detectors: neurons sensitiv for
- frequ. modulation
- interaural amplitude diff.
- interaural phase diff.

3. interactions: nur auf Verhaltensebene:
a. Sekuler
b. Sophie Wuerger
c. Ehrenstein: cross-modal aftereffect: auditory displacements after adapt. to visual motion

========

In our case, however, it is evident from the mere structure of the data that they cannot be explained by some version of probability summation of independent sensory channels. They require the postulation of a genuine (neural) summation at a relatively low level because the basic shape of the two-dimensional visual-auditory discrimination thresholds is logicaly inconsistent with probability summation in two important respects. First, the observed improvements for bimodal stimuli cannot be explained by a two-dimensional psychometric

function, which is consistent with probability summation, given the typical shapes of the intramodal psychometric functions. Second, and this is an even more interesting fact, which could not be revealed by a detection task but only with our discrimination paradigm: The observed summation is highly selective. Only those auditory-visual stimulus combinations, which are ecologically relevant, are directly integrated, whereas the other combinations came closer to a statistical pooling by probability summation. These combination-specific interactions cannot be the result of a simple bias, which is prevented by the forced choice design, and they are logically inconsistent with a probability summation of the outputs of independent sensory channels. Since we can definitely perceive increments as well as decrements in both modalities, a statistical summation can only yield the same positive pooling effects, irrespectice of the signs of the component changes, and is thus clearly inconsistent with the present results.

==========

In conclusion, we have presented the novel paradigm of bimodal discrimination threshold curves for the investigation of visual-auditory interaction effects. This discrimination paradigm closes a gap in the area of multisensory research, as compared to its long and succesful application in classical unimodal research (Graham, 1989; Zwicker, Fastl and Frater, 1999), and it offers special opportunities with respect to the distinction of “statistical” and “neural” interaction effects. Furthermore, it enabled a comparison of ecological valid vs. non-valid multisensory stimulus configurations within a uniform experimental setting. Our results give a clear indication for a “true” or “neural” summation of the information from the two modalities at a low to intermediate processing level, as opposed to a mere statistical pooling effect on a higher decision stage. Furthermore, they show that this summation is specific for ecologically valid stimulus combinations, which in this case correspond to the relative motion between a sound-emitting object and the observer, a situation with high behavioral significance. The observed effects can thus be regarded as prototypical example for the advantages of a true combination of informations from different sensory channels on an relatively early level in the neural processing hierarchy.

==========