Integration of sensory information within touch and across modalities

Marc O. Ernst, Jean-Pierre Bresciani, Knut Drewing & Heinrich H. Bülthoff
Max Planck Institute for Biological Cybernetics
72076 Tübingen, Germany
marc.ernst@tuebingen.mpg.de

***Abstract*—We perceive the world surrounding us via multiple sensory modalities, including touch, vision and audition. The information derived from all these different modalities has to converge in order to form a coherent and robust percept of the world. Here, we review a model (the MLE model) that in the statistical sense describes an optimal integration mechanism. The benefit from integrating sensory information comes from a reduction in variance of the final perceptual estimate. We here illustrate this integration mechanism in the human brain with two examples: the fist example demonstrates the integration of force and position cues to shape within haptic perception; the second example highlights multimodal perception and shows that tactile and auditory information for temporal perception interacts in a way predicted by the MLE integration model.**

Haptic Perception; Multimodal Interactions; Cue Integration; Maximum-Likelihood Estimation.

I. Introduction

A. *Integration of Sensory Information*

Our brain is constantly fed with a continuous stream of sensory information acquired from all the different sensory modalities. The resulting multimodal percept has to be coherent and unambiguous to enable interactions with the environment. But how does the brain come up with such a unique percept? To illustrate the problem, imagine driving a nail into wood using a hammer. The position of the nail in space can be seen, but may also be felt, while holding the nail in one hand. That is, vision and touch now provide redundant information about the nails position in space. To accurately hit the nail with the hammer the position information from the two modalities must be integrated into one common representation. Slight discrepancies in the representation of information between the modalities, as they naturally arise due to the probabilistic nature of sensory estimation (cf. next section), results in an interesting situation: the observer either has to decide which information (modality) to trust in a given situation (vision or touch) or it has to find a way to best combine the discrepant information to come to an optimal decision (or action).

On the other hand, having more than one (redundant) estimate available can be an advantage: The accuracy with which an environmental property can be judged increases with the number of individual perceptual estimates available. In the hammer example, the position in space can be estimated more reliably using both modalities (vision and touch) instead of only one. One could speculate that this may be one reason why it is better for you to hold the nail yourself, instead of having someone else hold it for you while hiding it with the hammer.

Not all information derived from different modalities is redundant. In the majority of cases information derived from the different modalities will be complementary in nature, such as when feeling an object's weight while seeing its color. Naturally, different combination rules have to be applied for combining such complementary information into a stable percept (cf. [1] for a recent review). Here, we concentrate on the integration mechanisms for redundant sensory information, such as that for a spatial position, that can be seen and felt, that for haptic shape that can be derived from force and position cues (see Chapter II), or that for a number of sequentially perceived events that can be seen, felt, or heard (see Chapter III).

B. *The probabilistic nature of sensory estimation*

The problem of sensory combination can be understood using signal detection theory [2]. Perception is a probabilistic process. If one estimates some environmental property, such as an object's size, this estimate will have a variance associated with it. As a result, if the same environmental property is estimated consecutively 100 times, all 100 perceptual estimates may vary slightly. Figure 1 shows schematically the probability density function for the estimation of an object's size s. In the simplest case this probability density function has a Gaussian shape and is unbiased. This curve is then defined by its mean $\bar{S}$, which for an unbiased estimator corresponds to the objects size, and its standard deviation σ:

$$\hat{S} = N(\bar{S}, \sigma).$$

If we define the reliability r as the inverse of the variance σ^2:

$$r = 1/\sigma^2,$$

then the larger the variance (or standard deviation) of the associated distribution the less reliably is the associated perceptual estimate.

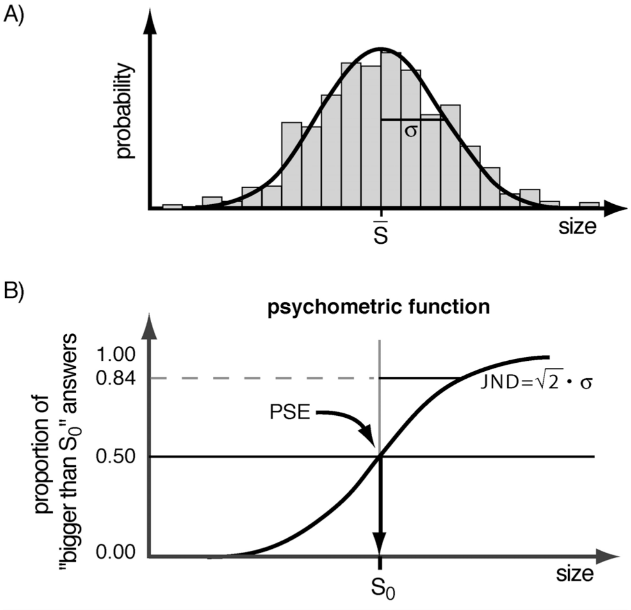

Figure 1. A) Schematic illustration of the probability density function for the estimation of an object's size s. The histogram indicates the distribution of answers derived from the size estimation process. The fitted curve has a Gaussian shape (with standard deviation σ and mean $\bar{S}$) and indicates the probability density function. B) Schematic drawing of a psychometric function derived using a 2-interval forced-choice task given the probability density function for estimating the object's size from A. The just noticeable difference (JND) derived at the 0.84 point corresponds to $JND = \sqrt{2} \cdot \sigma$. PSE is the point of subjective equality.

How can the variance or the reliability of a sensory signal be estimated experimentally? For this, classical psychophysical discrimination paradigms such as a 2-interval forced-choice (2-IFC) task can be used. Using the 2-IFC task subjects have to compare, for example, the sizes of two objects presented sequentially (a standard object of size S_0 and a comparison stimulus of size S). If the difference in size between the two intervals (S-S_0) is large, subjects will have no problem discriminating them, and consequently they will make only a few errors. With decreasing size difference however the error rate will rise. If the probability density functions for S and S_0 are Gaussian with identical variance σ^2, the resulting psychometric function is a cumulative Gaussian (see Fig. 1). The "Just Noticeable Difference" (*JND*) defined at the 84% level (the difference between the 50% and the 84% points) provides an estimate

$$JND = \sqrt{2}\sigma$$

for the variability of the underlying Gaussian distribution.

C. The MLE model for sensory integration

"Redundant signals" may to some degree sound like a waste of information. But actually this is not necessarily so. There are two major advantages in having available redundant information: the first is that the system is more robust, because when there is one estimate not available at a given time (or its information is degraded) the other estimate can substitute for it. The second advantage is that the final estimate becomes potentially more reliable compared with the reliability of the individual estimates feeding the combined percept.

What would be the statistically optimal strategy for combining redundant sensory information? Figure 2 shows the probability density functions for two independent estimates derived from two different modalities. In the example discussed here it is a size estimate that is derived from the visual and haptic modalities ($\hat{S}_V$ and $\hat{S}_H$). According to the "Maximum-Likelihood-Estimation" (MLE) scheme the integrated size estimate $\hat{S}_{VH}$ is a weighted average across the individual sensory signals with weights w_i that sum up to unity (the indices *i*, *j* refer to the individual modalities) [3].

$$\hat{S} = \sum_i w_i \hat{S}_i \quad \text{with} \quad \sum_i w_i = 1\cdot$$

Optimally, weights are chosen to be proportional to the reliability of a given signal. That is, if the visual modality provides the more reliable information in a given situation this signal is weighted higher in combination.

$$w_j = \frac{r_j}{\sum_{i=1...,j,...N} r_i}.$$

In the example shown in Figure 2 the variance associated with the visual estimate is less than the variance associated with the haptic estimate. That is, the visual information is more reliable. Therefore, the combined estimate being the weighted sum is closer to the visual than the haptic estimate. Under other circumstances where the haptic modality provides the more reliable estimate the situation is reversed.

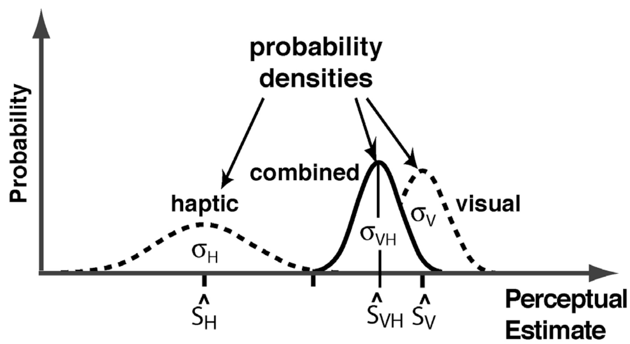

Figure 2. Schematic drawing of the probability density functions of the individual visual and haptic estimates and of the combined visual-haptic estimate, which is a weighted average according to the MLE integration rule. The variance associated with the visual-haptic distribution is less than either of the two individual estimates.

The variance of the combined estimate will be less then that of either of both individual estimates feeding the combination process. In other words the reliability can only improve by combining information (provided the signals and their noise distributions are independent, so there is no interaction). Following the MLE principle the reliability of the combined estimate is the sum of the reliabilities of the individual estimates:

$$r = \sum_i r_i$$

One can show that the MLE integration scheme is statistically optimal in that it provides the most reliable unbiased sensory estimate, given that the individual estimates are Gaussian distributed and that these noise distributions are independent. However, even if the noise distributions of the individual estimates show some correlation one can still benefit from combining sensory information and the combined estimate may become more reliable than the individual estimates [4].

In a recent study we could show that humans actually integrate visual and haptic size information in such a statistically optimal fashion [5]. Others now showed that this finding of optimality holds not only for the integration across vision and touch, but also for the integration of information across and within other modalities, such as audition or vision [6, 7, 8]. This suggests that maximum-likelihood-estimation is an effective and widely used strategy exploited by the perceptual system. In the following two chapters we illustrate some of the mechanisms predicted by the MLE integration scheme.

II. Within-haptic Cue Integration

A. Cues to shape in active touch

So far, the MLE model on cue integration has been well supported for within-visual and for crossmodal integration. Within the haptic modality, however, the integration of different cues is yet hardly examined in a corresponding systematic fashion. Whether the MLE model generalize to within-haptic integration is particularly interesting for the case of active touch. This is because during active touch the observer is able to actively pick up the information that is most important for him. Clearly this distinguishes active touch from the less dynamic situations so far examined where the observer integrates the stimulation that he/she obtains mostly passively.

This series of experiments aims at systematically exploring cue integration during active touch. For this, a recent finding from haptic curvature perception provides a good starting point. Robles-de-la-Torre and Hayward [9] distinguished between positional and force cues in the perception of small-scale bumps (amplitude 3 mm): When sliding a finger across a bump on a surface, the finger follows the geometry of the bump providing positional cues for the shape. At the same time the finger is exposed to forces related to the slope of the bump (force cues). In a psychophysical experiment, participants moved a probe over a physical bump, hole or plane; a custom-made device allowed to manipulate the forces opposing/supporting the movement. Nearly all participants reported to feel the shape indicated by the force cues and not by the positional cues. This finding tends to indicate absolute dominance of force over position cues.

However, the task in the experiment described by Robles-de-la-Torre and Hayward was to categorize perceived objects in bumps or holes, so that the observed dominance may just indicate a relative dominance of one cue over the other which manifests itself as an absolute dominance in the classification responses given. This is in contrast to an absolute perceptual dominance of force over position cues.

In Experiment 1 we independently and systematically varied force and position cues to haptically perceived curvature of 3D-arches and quantified the percept by psychophysical methods. For that purpose, we constructed a set of virtual arches with conflicting cues, where we intermixed force and position cues related to curvatures between 0 and 16 /m. Participants compared these conflicting-cue shapes to shapes with cues consistent.

B. Experiment 1 – Integrating force and position cues

1) Method

The haptic stimuli were virtual arches of 30 mm width and 50 mm depth that were curved in height along the depth axis. A PHANToM 1.5 haptic device (SensAble Technologies) displayed the stimuli to the right index finger of the participants. This device displays virtual objects by measuring the position of the finger and, then, exerting appropriate reaction forces on it. We defined the positional cues of curvature by the positions of the onsets of the object's reaction forces and the force cues as the directions of these reaction forces. In natural objects reaction forces are normal to the surface and, thus, force and position cues are highly correlated. However, with the PHANToM device we were able to disentangle these cues and combine force and position cues from different arches (Fig. 3).

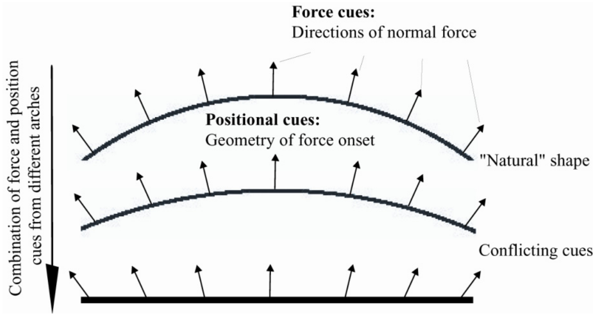

Figure 3. Force and position cues in arches and their combination

We constructed nine standard stimuli in which we completely intermixed force and position cues to curvatures of 0, 8, and 16 /m. To each standard curve belonged a set of 13 "non-conflicting" comparison curves, in each of which force and position cues indicated the same curvature. Curvatures of the comparison stimuli were distributed in a range of +/-9/m around the expected point of subjective equality (PSE) for the corresponding standard curve, i.e. around that comparison curve which we expected to be indistinguishable from the standard curve.

This work was supported by the Max Planck Society and by the 5th Framework IST Program of the EU (IST-2001-38040, TOUCH-HapSys).

Using a 2-IFC task each trial in the experiment consisted of the sequential presentation of a standard and a comparison. Then, participants had to decide which of the two stimuli was more convex. We displayed each pair of stimuli 16 times to each participant following the method of constant stimuli. The order of stimuli within the trials was balanced, the order of the pairs completely randomized. The entire experiment lasted about 8 hours that were divided into four sessions accomplished on different days. Seven right-handers took part in the experiment - naïve to the purpose of the study and paid to participate.

In the data analysis we determined for each participant and each standard stimulus the psychometric function, that is the proportion of trials in which the comparison was perceived as more convex than the standard against the curvature of the comparison. Fits were done with a cumulative Gaussian function using the psignifit toolbox version 2.5.41 for Matlab which implements the maximum-likelihood fitting method described in [10]. From these fits, we obtained individual PSEs.

2) Results & Discussion

Individual PSEs (Fig. 4) entered into an ANOVA with the within-participant variables Position Curvature (0, 8, 16/m) and Force Curvature (0, 8, 16 /m). Both main effects were reliable, (Position Curvature, $F(2,12)=228.8$, $p<.001$; Force Curvature, $F(2,12)=80.4$, $p<.001$), indicating that PSEs systematically increased with increasing curvature specified by either cue. So, both force and position cues contributed to perceived curvature, which is consistent with the MLE model.

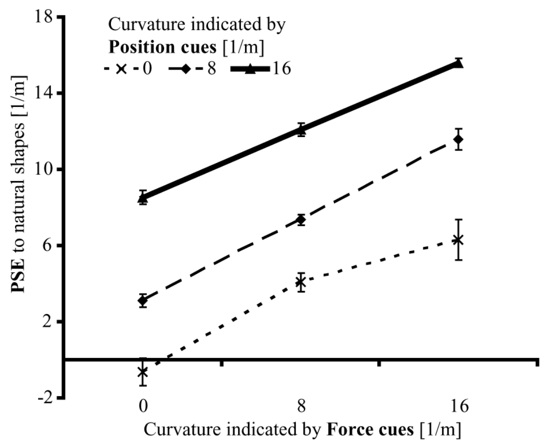

Figure 4. Average PSEs and their standard errors in Experiment 1

The MLE model further predicts that the combination of two cues can be described by a weighted linear combination. Here, we found a marginally significant interaction between Force Curvature and Position Curvature, $F(4,24)=3.6$, $p<.05$. This points to some violation of linearity. However, the linearity assumption is only true if the reliability of the signal does not change with curvature. Further, a multiple linear regression (intercept = 0) of the average PSEs on the curvatures indicated by position and force cues explained the variance between PSEs to 99%, indicating that weighted averaging is at least a good approximation. The relative weight for the force cue was 46% and for the position cue 54%. Finally, individual multiple regressions confirm that each person combines force and position cues approximately linearly (variance explained > 94%) and that, also individually, both cues (force weights: 28% to 61%) contribute to perceived curvature. Being questioned after the experiment none of the participants reported to have noticed the conflicts between the force and position cues.

Taken together, findings from this experiment suggest that haptically perceived curvature is a weighted average of both positional and force cues. This is consistent with the MLE model on cue integration. Another prediction of the MLE model is that the cue weightings depend on the relative reliability of the respective cue. The marginal interaction we found between the two cues may hint at a change of weight with curvature (i.e., the size of the aches) Experiment 2 provided a first test of this prediction. In Experiment 2, two sets of virtual shapes realized shallow vs. high arches. Keeping the length of the curved section constant, we intermixed force and position cues indicating curvatures of 0, 5, and 10/m (shallow arches) and of 20, 25, and 30/m (high arches), respectively. We expected that curvature perception in high arches position cues have relatively more weight than in the perception of shallow arches. We expect this, because positional changes of the fingers during stroking high arches are much more pronounced than for shallow ones.

C. *Experiment 2 – Changing weight with size of arch*

1) Methods

Apparatus, stimulus construction and the procedure in single trials were the same as in Experiment 1. There were three within-participant variables: Arch Rise (*shallow* [x = 5/m], and *high* [x=20/m]), Force Curvature (x-5, x, x+5/m), and Position Curvature (x-5, x, x+5/m) realized in 18 standard shapes.

Each participant conducted one double-staircase (1-up/1-down) per standard stimulus (random order), in each of which two adaptive staircases were interleaved. Step size was adjusted from 8/m at the beginning to 2/m for shallow arches and from 6/m to 1.5/m for high arches within 3 reversals. Each staircase stopped after 8 non-reducing reversals. From averages of the comparisons' curvatures across these 8 reversal points, we estimated the PSEs [11]. The experiment lasted about 2.5 hours including an initial short practice phase. We tested 14 paid, naïve participants (all right-handed).

2) Results & Discussion

Individual PSEs (Fig. 5) entered an ANOVA with the variables Arch Rise, Force Curvature, and Position Curvature. A main effect of Arch Rise, $F(1,13)=3420.1$, $p<.001$ confirmed that the high arches were generally perceived as being more convex than the shallow arches. Main effects of Position Curvature, $F(2, 26)=357.1$, $p<.001$, and Force Curvature, $F(2, 26)=464.5$, $p<.001$, confirm the finding from the previous experiment that PSEs increase with increasing curvature specified by either cue. Most importantly, interactions of Arch Rise with Position Curvature, $F(2, 26)=26.2$, $p<.01$, as well as with

Force Curvature, $F(2, 26)=39.2$, $p<.01$, indicate that the two cues contributed differently to perceived curvature in the two sets of arches. In other words, the relative weighting of the two cues changes with the size of the arch. There were no other reliable effects (Fs<1).

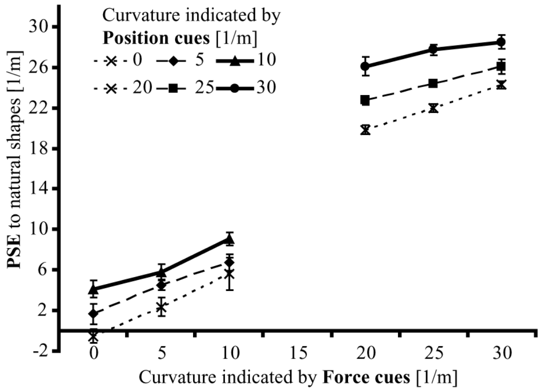

Figure 5. Average PSEs and their standard errors in Experiment 2

In a second step, we calculated multiple regressions (intercept = 0) of the PSEs on the standards' force and position curvatures separated by Arch Rise: For shallow arches the relative force cue weight was 67%, the position cue weight 33% and variance was explained to 99%; for high arches the values were 41% force, 59% position, and 99% explained variance. Consistent with the MLE model, the high amount of explained variance shows for each arch set alone, that perceived curvature can be described by weighted averaging of the curvatures indicated by the two cues. Cue weights show that force cues influence the percept more in the set of shallow as compared to high arches and vice versa for position cues. This is consistent with our hypothesis that the more pronounced the finger amplitudes the higher the relative position cue weight. Also consistent is the relative force cue dominance in the shallow arch set that was previously observed [9]. Most importantly, the difference in cue weights depending on arch set is consistent with the prediction of the MLE model that cue weightings change with the relative reliabilities of the respective cues. Again, none of the observers reported to have noticed the conflicts between the cues.

D. *Conclusion: Within haptic cue integration*

In this series of experiments we have explored whether principles of cue integration known from integration within and across other modalities and formulated by the MLE model can be extended to active haptic perception. Experiment 1 demonstrated that the integration of force and position cues to haptic shape can be described by a linear weighting model. Most importantly, this was true for each individual subjects data. Experiment 2 corroborated this assumption and, further, demonstrated that cue weights depend on the curvature of the shape. Position cue weights were higher for curvature perception of high as compared to shallow arches - probably mediated via changing cue reliability. Taken together, also within-haptic cue integration seems to be well described by the MLE model. Thereby, to our knowledge the present series of experiments is the first systematic study on cue integration during active haptic exploration.

III. AUDITORY-TACTILE INTEGRATION

A. *Combining multimodal temporal events*

Our everyday interactions with the environment provide us with a continuous stimulation of our different sensory channels. The central nervous system (CNS) has thus to deal with a pool of multimodal signals providing information of different nature concerning body/ environment relationship. In many cases, the occurrence of a specific value of the signal in one sensory modality is accompanied by a "corresponding" specific signal in one or more other modalities. For instance, when knocking on a door, one gets congruent visual, tactile and auditory feedback, this feedback being specific to the characteristics of the action (e.g., number of times one knocked, delay in between two knocks, knocking force's intensity). Several psychophysical experiments suggested that these redundant sensory signals are automatically co-registered to derive a coherent unified percept of the presented stimuli [12-18]. One convenient way of determining whether two signals are automatically combined consists in testing whether a to-be-ignored background signal can bias the perception of a to-be-assessed focal signal. In line with this, we tested whether auditory and tactile signals are automatically combined for tactile taps (Experiment 3) and auditory beeps (Experiment 4) perception.

B. *Experiment 3 – Effect of beeps on taps perception*

1) *Methods*

In the first experiment, we tested whether the perception of tactile sequences of taps (2 to 4) delivered on the index fingertip can be modulated by simultaneously presented sequences of 'to-be-ignored' auditory beeps when the number of beeps differs (less or more) from the number of taps. Four auditory conditions were associated to the presentation of the tactile sequences: 'No Beep' (baseline performance for tactile perception), 'One Beep Less' (# beeps = # taps-1), 'Same Amount' (# beeps = # taps), and 'One Beep More' (# beeps = # taps+1).

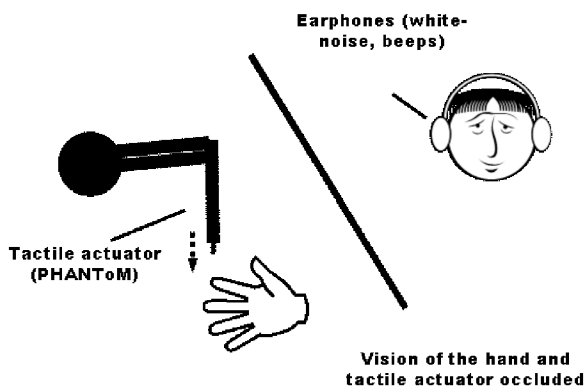

Figure 6. Experimental set-up for tactile auditory stimulation.

The experimental set-up is schematically represented in Figure 6. The tactile taps were again generated via a PHANToM force-feedback device. A metallic pin of 1 mm in diameter indented subjects' skin by approximately 2 mm with a force of 1 N. The subjects could not see their hand or the force-feedback device.

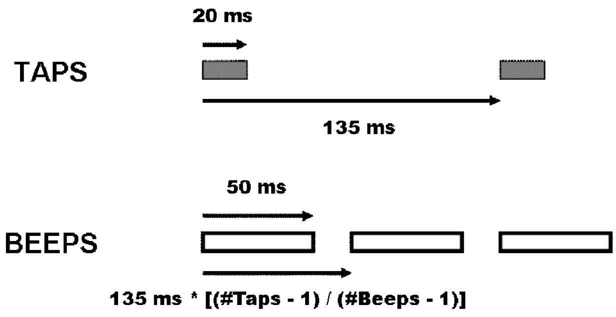

Figure 7. Temporal sequence of haptic and auditory events. Haptic taps were 20 ms lang with ans ISI of 135 ms. Beeps were 50 ms long with variable ISI (see formula in figure).

Each tap lasted 20 ms and the delay between the onsets of two successive taps was 135 ms (see Figure 7). The beeps were auditory tones (790 Hz, 74 dB) delivered via earphones. Each beep lasted 50 ms and the delay separating the onsets of two successive beeps varied so that the onsets of the first and last beeps coincided with the respective onsets of the first and last taps. For the whole duration of the experiment, the earphones also emitted a white-noise (71 dB) to mask any external auditory disturbance.

2) Results

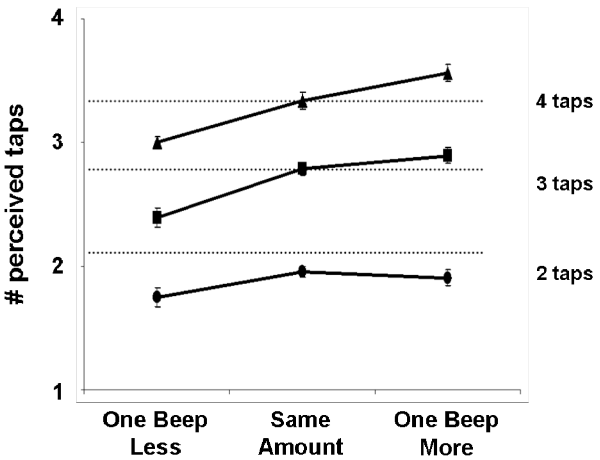

Figure 8. Number of perceived taps as a function of both the actual number of delivered taps and the auditory condition. The dotted lines represent subjects' average perception in the 'No Beep' condition

As shown in Figure 8, the perceived number of tactile taps was significantly (ANOVA) influenced by the simultaneous presentation of to-be-ignored auditory beeps. Indeed, the perceived number of taps not only depended on the actual number of delivered taps [$F(2, 46) = 521.11$, $p < .001$], but also on the number of simultaneously presented tactile taps [$F(3, 69) = 43.66$, $p < .001$]. Post hoc comparisons using a Sidak adjustment for multiple comparisons ($p < 0.05$) were then performed. Concerning the main effect of the number of taps, each of the 2 (mean = 1.93), 3 (mean = 2.71) and 4 taps (mean = 3.31) conditions were significantly different from each other. Concerning the main effect of the number of beeps, the 'No Beep' (mean = 2.74) and the 'Same Amount' (mean = 2.69) conditions did not differ from one another. The perceived number of taps in the 'One Beep Less' condition (mean = 2.38) was significantly lower than in any other auditory condition, whereas the perceived number of taps in the 'One Beep More' condition (mean = 2.79) was significantly higher than all but the 'No Beep' condition.

C. Experiment 4 – Effect of taps on beeps perception

1) Methods

In the second experiment, we tested whether the automatic combination of tactile and auditory signals also works the other way around, i.e., whether the perception of auditory sequences of beeps (2 to 4) can be modulated by simultaneously presented sequences of 'to-be-ignored' tactile taps when the number of taps differs (less or more) from the number of beeps. Four tactile conditions were associated to the presentation of the tactile sequences: 'No Tap' (baseline performance for auditory perception), 'One Tap Less' (# taps = # beeps-1), 'Same Amount' (# taps = # beeps), and 'One Tap More' (# taps = # beeps+1).

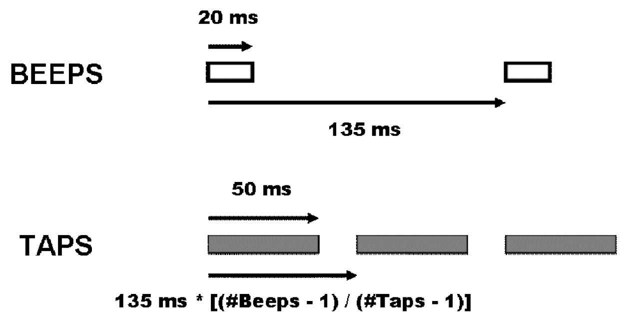

Figure 9. Temporal sequence of haptic and auditory events used in Exp. 4.

Each beep lasted 20 ms and the delay between the onsets of two successive beeps was 135 ms (see Figure 9). Each tap lasted 50 ms and the delay separating the onsets of two successive taps varied so that the onsets of the first and last taps coincided with the respective onsets of the first and last beeps.

2) Results

The results are presented in the Figure 10. The perceived number of beeps depended on the actual number of delivered beeps [$F(2, 46) = 834.40$, $p < .001$], but was also influenced by the number of simultaneously presented tactile taps [$F(3, 69) = 6.04$, $p < .01$]. Post hoc comparisons using a Sidak adjustment for multiple comparisons ($p < 0.05$) revealed that each of the 2 (mean = 1.85), 3 (mean = 2.93) and 4 beeps (mean = 3.75) conditions were significantly different from each other. However, concerning the main effect of the number of taps, only the 'No Tap' (mean = 2.79) and the 'Same Amount' (mean =

2.91) conditions significantly differed from one another. Neither the ‘One Tap Less’ (mean = 2.80) nor the ‘One Tap More’ (mean = 2.87) tactile condition did differ from any of the other conditions.

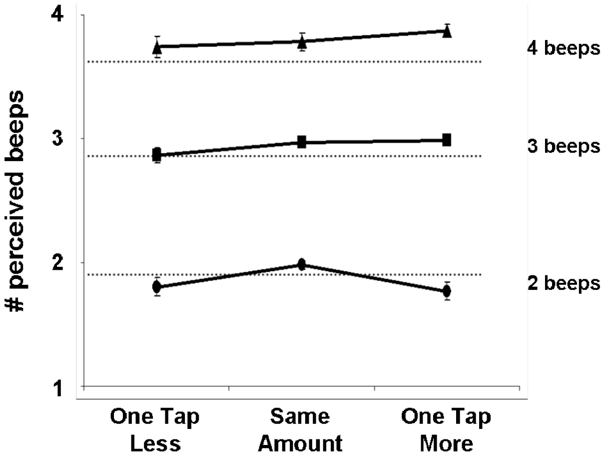

Figure 10. Number of perceived beeps as a function of both the actual number of delivered beeps and the tactile condition. The Dotted lines represent subjects’ average perception in the ‘No Tap’ condition

D. *Discussion*

The results of these experiments show that auditory and tactile signals are automatically combined both for tactile taps and auditory beeps perception, and this even when the subjects are explicitly instructed to ignore the “background” signal. This automatic combination likely results from the fact that the matching between co-occurring multimodal signals is very consistent across our everyday experience, so that the CNS can learn to co-register sets of redundant sensory signals and identify every single set as elicited by the same unique event or stimulus. Because multimodal cues reduce the variance of perceptual estimates [5, 19] and enhance stimulus detection [20, 21], such automatic combination of redundant-like sensory signals can be conceived as an optimization process.

Our results also show that the combination of auditory and tactile signals does not lead to similar results depending on whether the primary task is tactile or auditory. Indeed, if auditory background signals significantly biased tactile perception in the first experiment, no real bias of auditory perception could be evoked by tactile signals in the second experiment (no effect of the ‘One Tap Less’ and ‘One Tap More’ conditions). In this latter case, the only significant effect of the taps on auditory perception was an improved accuracy of the subjects when redundant auditory and tactile signals were available (‘Same Amount’ tactile condition) as compared to when only auditory signals were provided (‘No Tap’ tactile condition). This indicates that in this type of non-spatial task (i.e., counting the amount of events in given sequences), there is a relative domination of auditory signals on tactile signals. In this respect, our results are in line with previous experimental results reporting auditory-evoked biases of tactile perception in the non-spatial domain [16-18].

IV. Summary and Conclusion

This manuscript started out with presenting a model for multimodal sensory integration. This model – the MLE model – describes an integration mechanism that in the statistical sense is optimal. The criterion for optimality is defined as the precision of the system that is here determined by the reliability (i.e., the inverse variance) of the perceptual estimate.

We further illustrated details of the model using two sets of experiments. The purpose of the first set of experiments (Exp. 1 & 2) was to investigate whether cues to shape in active touch integrate according to the rules of the MLE model. The purpose of the second set of experiments (Exp. 3 & 4) was to investigate the integration of sound with tactile information for the perception of a sequence of short temporal events.

From the within-haptic cue integration experiments we concluded that the MLE model is suitable for describing the process of integration of position and force cues to haptic shape and that the weight given to each cue changes with the curvature presented – i.e., high arches have a higher position cue weight relative to small-scale arches, in which the force cue is weighted relatively higher. In practice this means that the percept of a curvature indicated by some positional profile can be influenced by the force profile provided along with the positional information.

In the future such interactions between force and position information may be exploited in the rendering of haptic objects and surfaces. For example, one could imagine that techniques that are well established in computer graphics – such as bump or normal mapping – can be transferred to the rendering of haptic objects in computer haptics. According to this idea the haptic bump or normal map would contain a 2D array of force vectors corresponding to the virtual normal vector of the surface to be represented. Since the force cue is the dominant component for small-scale structures, from our basic scientific studies presented here, we would predict that such “cheating” in form of a haptic bump or normal map would work as long as the haptic structures are relatively small (within the range of a few millimeters). I.e., we propose that on a small scale there is no need to have the positional information completely correct (according to the physical model) as long as the force profile is calculated appropriately. Most obviously the advantage of such a shortcut is the saving of compute power.

From the studies on auditory-tactile integration of sequences of short temporal events we could concluded that such multimodal information is automatically integrated, independent of whether attention is put on the one or the other modality. Such knowledge may now also be systematically exploited for the haptic rendering of objects and events. For example, one could imagine inducing the feeling of two haptic events by playing an accompanying sound but in reality there was only one haptic event present. Such “cheating“ may help simplifying the rendering complexity inherent in some haptic or multimodal applications.

Taken together, we demonstrated that a very general principle, such as the MLE model, can be used for explaining many phenomenon in haptic and multimodal integration. With this basic knowledge and its possibility to make systematic predictions, such phenomenon may now also find its way into applications of haptic or multimodal rendering and display technology.

References

[1] M. O. Ernst, and H. H. Bülthoff, "Merging the senses into a robust percept," Trends in Cogn. Sci. 8 (4), April 2004, pp. 162-169.

[2] D. M. Green and J. A. Swets, "Signal Detection Theory and Psychophysics," Peninsula Publishing, Los Altos, Ca, USA, 1988

[3] W. G. Cochran, "Problems arising in the analysis of a series of similar experiments," J. Royal Stat. Soc. (Sup.), 4, 1937, pp. 102-118.

[4] I. Oruç, T. M. Maloney, and M. S. Landy, "Weighted linear cue combination with possibly correlated error," Vis. Res. 43, 2003, pp. 2451-2468.

[5] M.O. Ernst, and M.S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," Nature 415, 2002, pp. 429-433.

[6] D. Alais, and D. Burr, "The ventriloquist effect results from near optimal crossmodal integration," Curr. Biol. 14 (3), Feb. 2004, pp 257-262.

[7] D. C. Knill, and J. A. Saunders, "Do humans optimally integrate stereo and texture information for judgments of surface slant?" Vis. Res. 43 (24), 2003, pp. 2539-2558.

[8] J. M. Hillis, S. J. Watt, M. S. Landy, and M. S. Banks, "Slant from texture and disparity cues: optimal cue combination," Journal of Vision , (*submitted*).

[9] G. Robles-de-la-Torre, and V. Hayward, "Force can overcome object geometry in the perception of shape through active touch," Nat. 412, 2001, pp. 445-448.

[10] F.A. Wichmann, and N.J. Hill, "The psychometric function: I. Fitting, sampling and goodness-of-fit," Perc. & Psych. 63 (8), 2001, pp. 1293-1313.

[11] J.C. Falmagne, "Psychophysical Measurement and Theory," in Handbook of perception and Human performance, vol. 2, K.R. Boff, L. Kaufman, and J.P. Thomas, Eds. New York: John Wiley & Sons, 1986, pp. 1.1–1.66.

[12] P. Bertelson, and M. Radeau, "Cross-modal bias and perceptual fusion with auditory-visual spatial discordance," Percept. Psychophys. 29, 1981, pp. 578-584.

[13] R. Fendrich, and P.M. Corballis, "The temporal cross-capture of audition and vision," Percept. Psychophys. 63, 2001, pp. 719-725.

[14] S. Morein-Zamir, S. Soto-Faraco, and A. Kingstone, "Auditory capture of vision: examining temporal ventriloquism," Cog. Brain Res. 17, 2003, pp. 154-163.

[15] L. Shams, Y. Kamitani, and S. Shimojo, " What you see is what you hear," Nature 408, 2000, p. 788.

[16] S. Guest, C. Catmur, D. Lloyd, and C. Spence, "Audiotactile interactions in roughness perception," Exp. Brain Res. 146, 2002, pp. 161-171.

[17] V. Jousmäki, and R. Hari, " Parchment-skin illusion: sound-biased touch," Curr. Biol. 8, 1998, p. 190.

[18] K. Hötting, and B. Röder, " Hearing cheats touch, but less in congenitally blind than in sighted individuals," Psychol. Sci. 15, 2004, pp. 60-64.

[19] W.C. Wu, C. Basdogan, and M.A. Srinivasan, "Visual, haptic, and bimodal perception of size and stiffness in virtual environments," P. ASME 67, 1999, pp. 19-26.

[20] S.C.A.M. Gielen, R.A. Schmidt, and P.J.M. van den Heuvel, "On the nature of intersensory facilitation of reaction time," Percept. Psychophys. 34, pp. 161-168, 1983.

[21] I.H. Bernstein, M.H. Clark, and B.A. Edelstein, "Effects of an auditory signal on visual reaction time," J. Exp. Psychol. 80, 1969, pp. 567-569.