It is also clear from our model that enhancement should be a function of modulation frequency. For a constant phase difference θ, time shift is related to the phase difference by $\Delta t = \theta / 2\pi f_m$ where f_m is the modulating frequency. When the modulation frequency increases, there is a corresponding decrease in tolerance of the AVgain to cross-modal asynchrony. By this argument, it is clear that at higher modulation frequencies the visual information requires more and more precise synchronization to effectively improve auditory processing. Thus there must be a frequency limit to the effect of visual enhancement.

Thus far little has been mentioned regarding the biological plausibility of the model presented here for audiovisual enhancement. Although this is not an issue that can be resolved definitely here, we do mention in passing that there are a number of ways in which the system in Figure 1 could have evolved as part of the neural processing mechanism. Here is one example: given that neurons tend to encode intensity through a logarithmic transformation, two neurons – one from the visual stream and the other from the auditory stream – both feed to a third neuron which encodes audiovisual information. To find the response of the third neuron, we add the outputs of the two input neurons. This is equivalent to multiplying the two input signals before taking the logarithm (i.e. log a + log b = log ab). While this argument is primitive at best, it does provide a starting point for further experimental studies.

In future work, we wish to refine our model and to make it made more physiologically accurate in several ways. One change would be the inclusion of auditory filters as part of the calculation of the detection statistic. Currently we have calculated enhancement at the carrier frequency only, but it is clear that detection occurs across a band of frequencies. Further experiments are also planned to explore the effect of reduced correlation between auditory and visual streams, and to use complex envelopes that contain multiple modulation frequencies. Such experiments would allow us to test our model more rigorously and to generalize to more complex audiovisual stimuli.

Acknowledgements

This work was sponsored by Defence Research and Development Canada and CIHR.

References

Erber, N. P. (1975). Auditory-visual Perception in Speech. *Journal of speech and hearing disorders. 40*: 481-492.

Fujisaki, W. (1), et al. (2004). Recalibration of Audiovisual Simultaneity. *Nature neuroscience 7.7*: 773-8.

Ghazanfar, A. A., & C. E. Schroeder (2006). Is Neocortex Essentially Multisensory? *Trends in cognitive sciences 10.6*: 278-85.

Kay, Steven M. (1998). *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. New Jersey: Prentice Hall.

Luu, S. and Wong, W. (2007). Crossmodal Enhancement: Investigating the Role of Timing in Audio-visual Enhancement. *Proceedings of ISP Fechner Day 2007*.

Meredith, M. A., and B. E. Stein (1986). Visual, Auditory, and Somatosensory Convergence on Cells in Superior Colliculus Results in Multisensory Integration. *Journal of neurophysiology 56.3*: 640-62.

SCALING CONFIDENCE CATEGORIES: EQUAL SPACING?

William M. Petrusic[1], Joel A. Lucas[1] & Joseph V. Baranski[2]

1. Department of Psychology, Carleton University, Ottawa, Ontario, Canada
2. Collaborative Performance and Learning Section, Defence Research and Development Canada – Toronto, Toronto, Ontario, Canada

Abstract

On each trial in a psychophysical comparison experiment, participants used the confidence categories "50", "60", "70", "80", "90", and "100 to indicate how certain they were that they had made a correct decision. We applied Case D of Torgerson's (1958) Law of Categorical Judgment (LCJ) to estimate the mean locations of the confidence category boundaries. Scale values for the confidence category boundaries were equally spaced on the underlying subjective probability scale and were identical in the speed and the accuracy stress conditions of the experiment. Excellent goodness of fit of the LCJ to the data was obtained in each condition.

Invariably when confidence ratings are taken as subjective probabilities, in studies examining how closely the confidence ratings correspond to the actual accuracy of the decisions rendered (see Baranski & Petrusic, 1994 for the definitional formula for the calibration index), it is tacitly assumed the objective confidence category labels represent equally spaced underlying numerical probabilities. It is not at all clear that this is the case.

Whether the actual confidence category labels can be taken at face value or not is an issue of some importance in how analyses are to proceed. For example, parametric analyses of mean confidence, as in ANOVAs, require the assumption that the confidence ratings are equally spaced and, in fact, can be viewed as a linear scale. Indeed, most notably, the tacit assumption is that the confidence category labels define points on an equally spaced scale of subjective probabilities. It is not at all well established that they do.

Typically, in probability assessment studies, confidence judgements, viewed as subjective probabilities, are rendered by participants selecting a value from the set, {50, 60, 70, 80, 90, 100}, with "50" denoting a guess and "100" complete certainty. Our approach to permitting a determination of the scale properties of such a set of confidence rating categories is to apply Torgerson's (1957) Law of Categorical Judgment (LCJ) to the matrix of frequencies of confidence category use associated with each of the stimuli in the experiment.

The LCJ posits scale values for both the stimulus items and the midpoints of the rating categories. Torgerson (1957) fully developed the most general forms of the LCJ permitting Gaussian variability in the representations of both the stimuli and the rating category boundaries as well as procedures for obtaining these scale values.

According to the most general form of the LCG, as developed by Torgerson (1958),

$$t_g - s_j = x_{jg}(\sigma_j^2 + \sigma_g^2 + 2r_{jg}\sigma_j\sigma_g)^{1/2}, \qquad (1)$$

with j=1, 2, …, n, and g=1, 2, …, m, and where m+1=number of categories, t_g =mean location of the g^{th} category boundary, s_j = mean location of the j^{th} stimulus, r_{jg} =correlation between momentary positions of stimulus j and category boundary g, and x_{jg} =unit normal deviate corresponding to the proportion of times stimulus j is sorted below boundary g.

However, invariably in applications, simplifying assumptions are made in order to obtain workable solutions (see Torgerson, 1958, pp207-210, Case D). Torgerson obtains least squares estimates of scale values for stimuli and category boundaries, upon assuming that the variance of the difference in discriminal dispersions (i.e., the term under the square root sign in Equation 1) is a constant. Under these conditions (see Torgerson, 1957, pp. 234-240) the LCJ for case D becomes

$$t_g - s_j = cx_{jg}\text{, with } j\text{=1, 2, …, } n \text{ and } g\text{=1, 2, …, } m. \qquad (2)$$

In the present experiment, the focus was on obtaining the t_g 's, the scale values for the confidence category boundaries, with a view toward determining the scale properties of the confidence category labels. In particular, the key question was whether the confidence categories can be viewed as equally spaced on the underlying confidence scale ranging from "50" to "100". In the present experiment, in the psychophysical tradition, comparisons of line lengths were required, under two conditions, one requiring sacrificing speed for accuracy, and in the other, a speed stress, at the expense of accuracy. Of interest, as well, was whether the differing demands for speed versus accuracy stress, might leave invariant the scale properties of the representations of the confidence categories.

Method

Participants. Twenty-eight Carleton University undergraduate students participated for one session lasting from 1.5-2 hours, in return for course credit.

Apparatus. The experiment was conducted on-line under the control of Turbo Pascal DOS based software running on a PC-IBM clone. Stimuli and instructions were presented on a Samsung SyncMaster 750s 18 inch video monitor.

Responses to stimuli were made on a keypad, positioned to the right of the monitor, with a faceplate (25 X 22 cm), slanted downward, with three rows of push buttons (2 X 2 cm each). The top row of six buttons was spaced horizontally to form a semi-circle and labelled from left to right (50, 60, 70, 80, 90, 100) to represent confidence categories. The middle row consisted of 2 buttons labelled "LEFT" and "RIGHT" which corresponded to left and right stimuli on the monitor which participants used to make discriminations of line-length. The single button of the bottom row, was used to initiate each trial and was also used to indicate an error during elicitation of confidence, and was labelled "START / ERROR".

Stimuli and Design. Stimuli consisted of pairs of white horizontal lines, presented on a black background. Stimulus pairs, denoted by combinations of pixel lengths, (x, y), where x represents the extent on the left and y on the right on the screen were: (100, 101), (101, 100), (100, 102), (102, 100), (100, 104), (104, 100), (100, 106), (106, 100), (100, 108), (108, 100), (100, 110), and (110, 100). Thus, the pairs were defined by six levels of difficulty with ratios of longer to shorter extent, 1.01, 1.02, 1.04, 1.06, 1.08, and 1.10, and two left-right presentation orders. On half the trials, participants selected the shorter line and the longer on the other half, thus resulting in 24 trials in a block. In each of two within participant conditions, each block was replicated 5 times for a total of 96 trials in each the two within participant conditions.

Ratings of confidence were obtained under two conditions. In the first, participants were instructed to sacrifice speed for accuracy. In the other condition, participants were instructed to sacrifice accuracy for speed in making their decisions within a 500 ms deadline. To try to ensure the relative demands for speed versus accuracy stress were met; payoffs and trial by trial feedback were provided. In the accuracy stress conditions, participants received $.02 for each correct response and were penalized $.02 for each error. In the speed-stress conditions, if participants met the deadline and responded correctly, they received $.02 each time. If they met the deadline but were incorrect, they received only $.01. If they did not meet the deadline but gave a correct answer they were penalized $.01. If they did not meet the deadline and were incorrect they were penalized $.02.

Procedure. Participants initiated each trial by pressing "START" on the key-pad, at which time they were presented with either the instruction "Longer" or "Shorter". One second later, a pair of lines was presented on the screen and remained on the screen until the participant responded. The next trial occurred two sec later. Participants were instructed to press the response key corresponding to the line length on the left or the right side of the screen corresponding to the instruction presented. Immediately after participants indicated a decision, the word "CONFIDENCE" appeared on the screen prompting participants to give a confidence rating by pressing a key on the keypad which best corresponded with their level of confidence (50-100). Participants were instructed that a rating of "50" indicated a pure guess or 50% chance of being correct, while a rating of "100" represented absolute certainty or a 100% chance of being correct and ratings from 60-90 represented varying levels of confidence between a pure guess and absolute certainty. Participants received feedback after each trial regarding the accuracy of their response and whether or not they met the deadline when it was required. After each response the message appeared "Primary Response Time Too Slow" or "Primary Response Time OK", and "Response Was Correct" or "Response Was Incorrect."

Results

The entries in Table 1 show that our participants used the confidence categories appropriately. In each condition, as the pairs become easier to discriminate (and discriminative accuracy increases) usage of the "100" ("Certain") category monotonically increases. On the other hand, usage of the "50" ("Guess") category monotonically decreases as the pairs become easier to discriminate (except for one reversal under speed stress). Thus, the data in the matrices provided below are entirely appropriate for the application of Torgersons' LCJ to obtain scale values for the confidence category boundaries.

Stimulus Pair	Confidence Category - Accuracy						Confidence Category - Speed					
	50	60	70	80	90	100	50	60	70	80	90	100
1.01	.243	.113	.097	.126	.112	.309	.245	.058	.099	.073	.117	.406
1.02	.175	.120	.136	.149	.108	.309	.252	.078	.086	.080	.083	.419
1.04	.134	.078	.126	.126	.130	.402	.204	.046	.075	.104	.101	.467
1.06	.073	.057	.073	.094	.138	.564	.177	.057	.089	.089	.067	.519
1.08	.039	.029	.052	.066	.130	.685	.148	.026	.052	.049	.107	.613
1.10	.010	.018	.029	.078	.099	.765	.114	.032	.032	.055	.076	.690

Table 1. Proportion of usage of each confidence category with each stimulus pair for the accuracy and speed stress conditions.

Stimulus Pair	Category Boundaries - Accuracy					Category Boundaries - Speed				
	50/60	60/70	70/80	80 /90	90/100	50/60	60/70	70/80	80/90	90/100
1.01	-0.697	-0.369	-0.118	0.199	0.498	-0.690.	-0.516	-0.248	-0.063	0.233
1.02	-0.934	-0.539	-0.174	0.202	0.490	-0.668	-0.439	-0.212	-0.010	0.199
1.04	-1.108	-0.799	-0.418	-0.090	0.238	-0.827	-0.674	-0.454	-0.179	0.752
1.06	-1.454	-1.126	-0.831	-0.533	-0.164	-0.927	-0.726	-0.459	-0.222	-0.053
1.08	-1.762	-1.491	-1.175	-0.893	-0.479	-1.045	-0.938	-0.752	-0.598	-0.300
1.10	-2.326	-1.911	-1.580	-1.103	-0.726	-1.205	-1.054	-0.923	-0.729	-0.499
t'_g	-1.380	-1.039	-0.716	-0.369	-0.024	-0.894	-0.724	-0.508	-0.300	-0.057

Table 2. Cell entries are standard normal deviates at each category boundary for each stimulus pair in the accuracy and speed conditions. Scale values for the confidence category boundaries are provided in the bottom row.

Least squares estimates of the t_g, scale values, were obtained using the procedure outlined in Torgerson (1958, p. 238). First, the relative frequencies were summated from left to right. Second, these summated relative frequencies were converted to the Gaussian transformed standard normal deviates shown in Table 2. Estimates of the t_g confidence category scale values were then obtained by computing the average of the normal deviates for each column (i.e., confidence category boundary). Scale values were then normalized to permit comparability between the accuracy and the speed stress conditions and to relate these values to the objective confidence category boundaries. The normalized scale values were placed on a scale ranging from "55" to "95". Figure 1 provides a view of these normalized scale values, for the accuracy and speed conditions, plotted at the mid-point of the objective confidence categories.

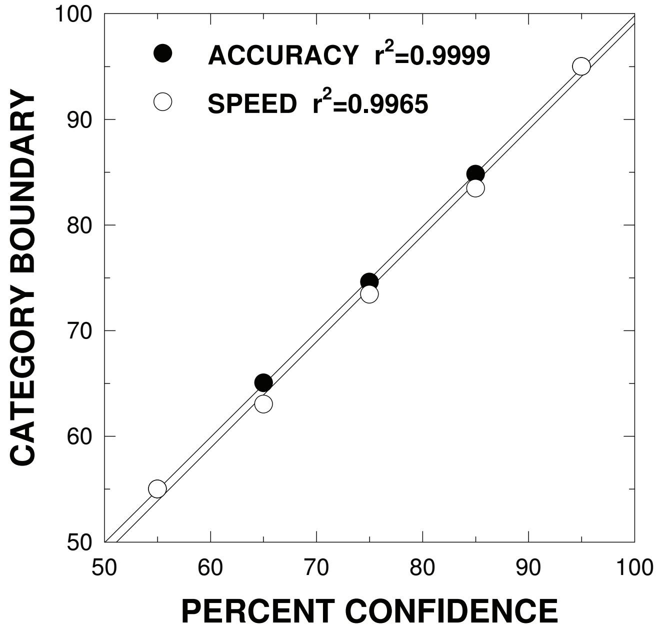

Figure 1. Normalized confidence category boundary scale values plotted against the mid-points of the objective confidence categories for the accuracy and speed stress conditions.

The plots in Figure 1 are clear in showing that the scale values for the confidence category boundaries are linearly related to the objective confidence categories. Importantly, the slopes of these plots are 0.998 and 1.004 for the accuracy and speed conditions, respectively. Thus, it is clear that the confidence category boundary scale values are equally spaced on the underlying subjective probability scale. Furthermore the demands for speed versus accuracy stress leave the subjective confidence scale invariant.

Goodness of fit

Tests of goodness of fit of the LCJ proceed as follows. The obtained estimates of the scale values for the category boundaries and the stimuli are used to derive the fitted proportion, p''_{jg}, which is the proportion corresponding to the standard normal deviate, x''_{jg}. For case D, $x''_{jg} = t'_g - s'_j$. Figure 2 provides plots of the fitted proportions plotted against the obtained proportions. As is evident, the fitted values correspond very closely to the obtained proportions for both the accuracy and the speed stress conditions. Thus, we conclude that the data are well fitted by Case D of Torgerson's LCJ.

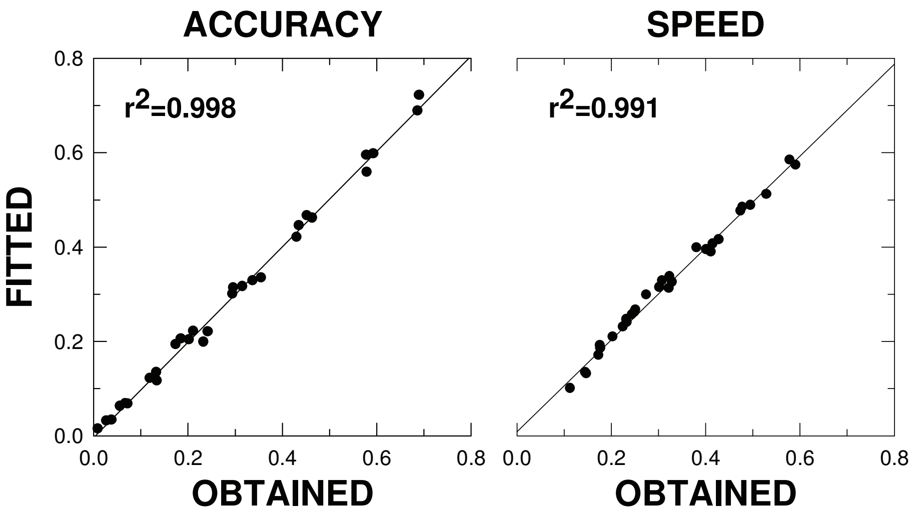

Figure 2. Fitted proportions, p''_{jg}'s plotted against the obtained proportions, p'_{jg}'s for the accuracy and speed stress conditions.

Summary and Conclusions

Application of Case D of Torgerson's (1958) Law of Categorical Judgment to the relative frequencies of confidence category usage in a line length discrimination task permitted estimates of the scale values for the confidence category boundaries. Upon normalization and rescaling, these scale values were linearly related to the objective confidence category boundaries in both the accuracy and speed stress conditions. The slope of each of these plots was essentially 1.0, thus showing that the confidence category scale values is equally spaced on the underlying subjective probability scale. Fitted proportions, based on the scale values for the category boundaries and the stimuli, corresponded very closely to the obtained, indicating that the Case D assumptions were justified.

References

Baranski, J.V. & Petrusic, W.M. (1994). The calibration and resolution of confidence in perceptual judgements. *Perception & Psychophysics*, **55**, 412-428.

Torgerson, W.S. (1958). *Theory and methods of scaling*. New York: Wiley.

Acknowledgement

This work was supported by a Natural Sciences and Engineering Research Council of Canada Discovery grant to William M. Petrusic.

THE TIME COURSE OF CONFIDENCE PROCESSING

Steven R. Carroll and William M. Petrusic
Department of Psychology, Carleton University, Ottawa, Canada
srcarrol@connect.carleton.ca; bill_petrusic@carleton.ca

Abstract

The time course of confidence processing is discussed in light of three decision-making experiments: two psychophysical tasks and one general knowledge task. Each experiment consisted of a confidence block of trials, wherein participants expressed a confidence rating following each rendered decision, a no confidence block, where confidence was never expressed, and a stop-confidence block, where participants were told to expect to have to express confidence but were instructed not to do so if a tone sounded. Stop-confidence tone onset delays were varied systematically. Decisional response time data reveal how confidence processing unfolds linearly throughout the primary decision-making process.

Baranski and Petrusic (1998) initiated the study of the locus and the time course of confidence processing in comparative judgements through investigation of the properties of the time to determine confidence. The time to determine confidence is defined by the time between the expression of the primary decision and the selection of a confidence category corresponding to the subjective probability that the decision rendered was correct. Baranski and Petrusic concluded that when the primary decision is made under speed stress confidence is computed post-decisionally. On the other hand, under accuracy stress, there was a strong suggestion that confidence is initiated and perhaps even completed during the primary decision process. Indeed, Petrusic and Baranski (2000, 2003) provided direct evidence that confidence is processed during the primary decision process under a stress for accuracy at the expense of speed. Using a between participants design in a line length discrimination task, one group of participants rendered confidence following the primary decision and the other simply made the comparative judgement. The requirement of confidence judgements substantially increased primary decisional response times (RTs). Moreover, confidence times varied systematically with confidence category, thereby implicating post-decisional confidence processing.

The Baranski and Petrusic (2003) and Petrusic and Baranski (2000, 2003) work demonstrate how rendering confidence increases primary decisional RTs but fails to specify the time course of the computation of confidence. For example, the balance of evidence may be computed during the final stages of decisional processing. On the other hand, the balance of evidence may evolve, continuously, over the course of decision processing. The presented experiments were designed to further examine both the loci and the time course of confidence processing by using a variant of Logan's (1983) "stop-signal" paradigm.

Experiments 1 and 2

Method

Participants. Forty-eight Carleton University undergraduate students participated in