

# DO YOU HEAR WHAT I HEAR? ADVANCES IN WEB-BASED PERCEPTUAL TESTING AND TRAINING

Dr. Richard Johnson, University of Alberta, rjohnson@bus.ualberta.ca  
Dr. Jane Saber, University of Texas at Tyler, jsaber@uttyler.edu

## ABSTRACT

*This paper describes a newly developed online perceptual testing and training tool called the Perceptual Chronograph. This tool was initially developed to test the reaction of salespeople to customer verbal and nonverbal cues in an experimentally designed 'sales transaction'. The Chronograph records the correct or incorrect identification of target or manipulated information, (signal detection theory), records the sensor's evaluation of the information, and also records reactions that a sensor would have in response to the identified information. The Chronograph has virtually unlimited potential for media richness: verbal, nonverbal, visual, contextual or even temporal information can be included. It also has the potential to be used as a training device, whereby exemplary sensors are tested and their response patterns analyzed to create a 'fuzzy gold' standard of behavior. Novice or less perceptually astute sensors (the 'novice') can be tested and their results analyzed. The differences between the exemplar and the novice can then be compared, either at the time of testing, or in a later training session. After feedback is given, the novice can be re-tested to ensure learning. Although the Chronograph was developed and tested in the sales context, it has learning and testing applications in many areas of research where a sensor (person or system) must perceive, evaluate, and respond to uncertain or conflicting information: (e.g., social perception, forecasting, and medical diagnosis).*

**Keywords:** communication, perceptual testing, signal detection theory, fuzzy gold standard

## INTRODUCTION

A salesperson is listening to her customer. She hears the customer say, "No, no, I'm not interested in purchasing this dress". At the same time, the customer is touching, examining, and holding the dress up to her body. What to do? The verbal cues indicated a lack of interest, yet the nonverbal cues indicate clear buying signals. Martha, the expert saleswoman, knows from experience that the nonverbal cues are more indicative of purchase intentions than the words. Martha believes the customer will buy the dress, attempts to close the sale, and sells the dress. Jennifer, on the other hand, has only started selling. She doesn't yet realize the importance of nonverbal cues and so when she hears the words "no, no", she assumes that the customer won't purchase; she does not try to close the sale and doesn't sell the dress.

How do you expeditiously train Jennifer, the novice salesperson, to be more like Martha, the expert? Further, how do you train other novices, faced with other types of ambiguous information, to behave like experts? Given that in 1995, 55.3 billion dollars was spent on sales training in the United States alone, tools developed to improve employee performance through training are critical. A newly developed testing tool, the *Perceptual Chronograph*, (PC), allows for the testing of novice 'information sensors', and has the potential to train these sensors to perceive, evaluate and react to ambiguous information in a manner similar to experts.

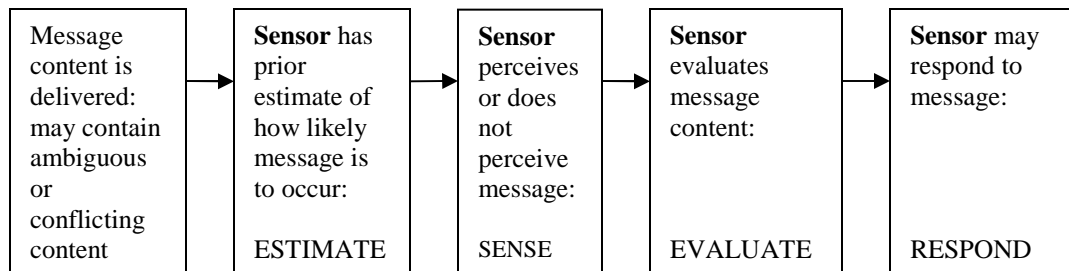
## **THEORETICAL DEVELOPMENT OF THE PERCEPTUAL CHRONOGRAPH**

The PC was initially developed in order to test the results of some commonly used sales effectiveness self report scales, as compared to the actual perception and reaction of salespeople [13, 10, 8]. Although the results of this research will not be reported here, [7] one of the main findings of that research was that salespeople behave in a manner that is not entirely consistent with their self-reports. This result was surprising, given the general acceptance of these scales; however, there had previously been no available method to empirically test the sensor's behaviour. The PC allows for such testing, and encompasses principles derived from the SESER model of communication [7] and signal detection theory [5] as found below.

The SESER model of communication was developed by Saber [7] in the context of sales communication. In this model, she outlines a number of steps that occur in the processing and use of potentially ambiguous or conflicting information by a sensor, as follows. First, before communication occurs, in most cases the 'communication-receiving' sensor will have a prior probability of the likelihood of a particular type of information being communicated in that particular context; (*estimation*). Second, the information is processed and perceived in the sensor's perceptual or sensory system, (*sense*). Third, the sensor will interpret the information's meaning (*evaluate*). Finally, in many cases, the sensor will respond to the ambiguous information (*respond*).

An example from the sales context is illustrative. A customer and her realtor are discussing a house purchase, and the customer makes a variety of relevant and irrelevant statements about the house. The realtor must determine which statements she should pay attention to, and which statements indicate that the customer is ready to 'write up the offer.' For example, the realtor must evaluate if the statement 'I want to consider buying this house' means that there is a 20, 70 or other percent chance that a customer will purchase. Generally, the sensor makes an evaluation (estimation) of this perceived information, which, in the case of salespeople, is the probability of purchase, (CPPA). This estimation will typically be influenced, on prior probabilities estimates (CPPP: estimations), based on prior experience or knowledge of occurrence probabilities of that category of information. For example, an excellent salesperson may regularly perceive information which indicates to him that a customer wants to purchase, and may estimate the likeliness of occurrence of this positive information as being quite high, (positivity bias / confirmation bias: 4, 2, 3). In contrast, a less successful or novice salesperson who rarely hears these customer indications of interest in purchasing may estimate the likeliness of this positive information occurrence as lower, (negativity bias: 1, 2, 3, 9). These prior probabilities are likely to have an impact on the evaluation of any actual information communicated.

Next, after perceiving and evaluating information, often some sensor action or response to the perceived information is expected. In the case of salespeople, there were four possible choices of behaviours: continue to listen, close the sale, end the transaction and respond to the customer. In the procedures, after listening to the statements and identifying a purchase probability, the salesman indicates which of these behaviours he would take. This allows a comparison of anticipated behaviour of novice and expert sensors, which can be analyzed using logistic regression procedures. An abbreviated graphical illustration of the SESER model is as follows:



**Figure 1.** SESER Model of Communication

The PC allows for the manipulation of the strength (ambiguous, non-ambiguous), valence (positive, negative) and presence or absence of relevant information, as well as the measurement and analysis of each of the sensor behaviors. Signal detection theory principles were also used to develop the PC as follows.

Signal Detection Theory (SDT) is a model which tests how ‘signals’ are detected (sensed) by a system or sensor in a background of interference, (‘noise’). SDT assumes that detection performance can be explained by two distributions: the presence or absence of a signal. SDT further assumes that there is an overlap between these so that a particular observation may have come from either distribution. The sensor has to decide whether the observation event came from a signal or noise distribution. Sensors can exhibit four states when exposed to signal/noise trials: **hit**: (H) sensors correctly identify that there is a signal, when there actually has been a signal,: P (Signal | signal), **false alarm**: (FA) sensors identify that there is a signal, when there was no actual signal: P (Signal | no signal), **correct rejection**: sensors identify that there was no signal, when there was no actual signal: P (No signal | no signal), and **miss**: sensors identify that there was no signal, when there actually was one: P (No signal | signal). The resulting identifications can be plotted on a two dimensional space as points in a plot of P (Signal | signal) versus P (Signal | no signal) since there are only two independent probabilities. If the sensor is perfect in her identifications of signal and noise, the hit rate will be 1, and the false alarm rate will be 0, [5].

SDT splits the variability in the data into two independent components: perceptual and decisional. The perceptual index,  $d'$ , is a measure of how well the sensor discriminates between noise and signals. The decision index,  $\beta$ , (bias) is the decision rule used by the sensor to arrive at his or her decisions: it is the tendency to favour one response over another, (type I or type II error). This rule could be conservative, where a low false alarm and hit probability would result, or lenient, which would generate very high false alarm and hit probabilities. Bias depends on the costs and values associated with making errors, as well as prior estimations of probabilities of the signal, [11]. An example of bias is helpful. When salespeople think the customer is ready to purchase, they will likely try to close the sale. If the customer is not ready to buy, and the salesperson tries to close anyhow, the customer may feel that the salesperson is ‘pushy and aggressive’ and walk out of the store. If it is likely that customers will leave the store if pushed, bias is likely to be conservative: salespeople won’t ‘push’ to sell unless there are strong indications that the customer wants to purchase. However, if the salesperson will never see the customer again and will not get referrals from that customer, the cost of trying to close when the customer is not ready to purchase is reduced. Here, a liberal bias score is more likely.

One of the major challenges in the application of SDT to ambiguous information contexts is that it is difficult to determine what an 'accurate' assessment of that information consists of. For example, how do we know that Martha, the expert saleswoman in the introductory example is 'right' in her assessment that the customer is likely to purchase the dress? Specifically, in the application of SDT to socially ambiguous cues, there is a need for an objective standard by which to judge accuracy of perception and indications of bias, [11]. Fortunately, the development 'fuzzy gold standard', or the cultural consensus model [14, 6] is designed to estimate the answers to a series of questions when the 'correct' answers are unknown. The model estimates individual raters' competencies, and generates estimates for the answers and the degree of confidence in each answer. Weller and Mann, [14] describe and outline the model, as follows. This model is based on three principles: 1. There is only one set of 'correct' answers, 2. Readers respond independently of one another, 3. All cases cover a single related topic at the same level of difficulty. The essence of this model is the notion that the agreement between two readers is a function of the degree to which each knows the correct answer. From the pattern of agreement between pairs of raters, individual knowledge or competency approximations can be estimated. Thus, by estimating expert or exemplar responses, a context-appropriate standard can be established.

In summary, measures of perceptual accuracy and bias can be tested by designing an experiment which varies the levels of noise, signal and signal strength, or a number of other variables. In the Saber [7] research, the experimental design was limited to the strength, valence and presence/absence of relevant verbal and paralinguistic information embedded in a number of noise statements, however visual, temporal and contextual information, in the form of streaming video, can be readily included. The procedures involved in the PC operations will be described below [7].

## **PROGRAM OPERATION**

The basic operations of this program, as used in the sales communication context will be described below. Because of the software flexibility of the PC, modifications to include different types of ambiguous information, (including different subject matters such as medical diagnosis, medical care, any other issues of social perception, or a multitude of other applications) would be relatively inconsequential. The procedures are as follows.

Sensors logged onto the testing website. They were given a unique password and test name, and signed on. They next saw a consent page which outlined the nature of the study, and if they agreed to participate, they pressed the "I have read the above and agree to participate" button and the program moved forward. If they declined, they pressed the "I decline to participate" button and the program went to a screen that said "Thank You for Your Time."

If the sensor agreed to participate, the sensors read a cover story where a fellow real estate agent friend had to go away on a family emergency, and had asked the sensor to show a house to a client on his or her behalf. Since it was an emergency, all that the friend had explained about the client was that it was the client's second visit to the house, and that if the client bought the house, the sensor would get a full commission on the sale. Thus, it was impossible for the sensor to have

any prior knowledge about the customer, effectively taking away any prior person schema that the sensor may have otherwise used.

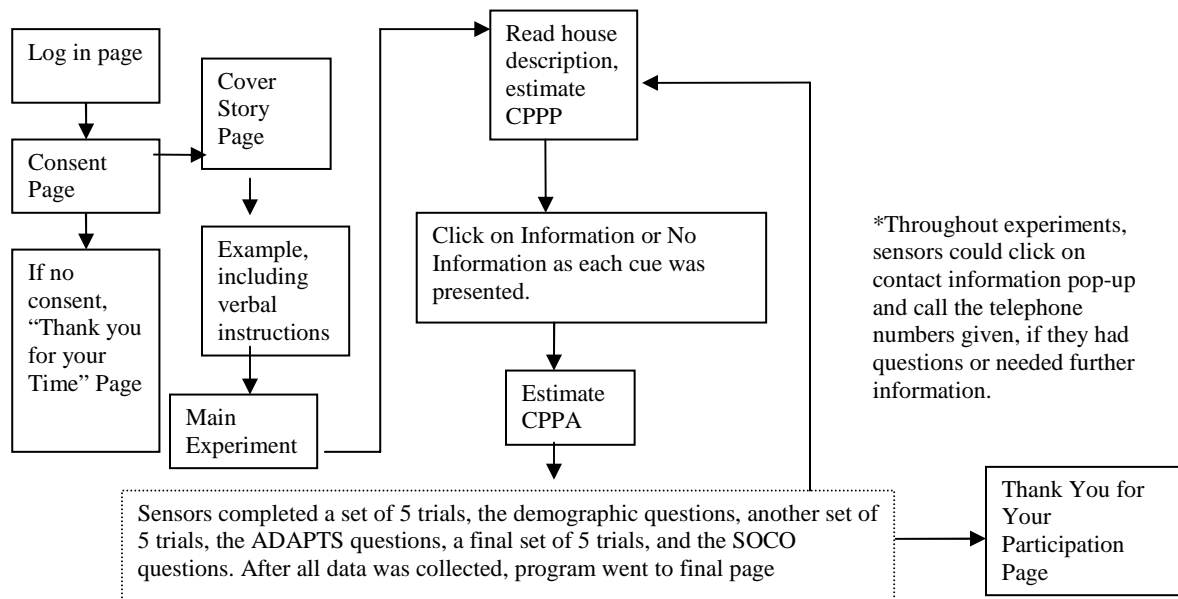
Next, the sensor read a house description. Based on the house description alone and the fact that this was the client's second visit, sensors were asked to make a customer purchase probability estimate, which is the prior probability of signal occurrence: CPPP. Once they estimated an initial CPPP, the sensor heard comments that the 'client' made while viewing the house, (hereinafter, "the conversation"). There were 15 trials in total, each presenting nine statements. Six of the statements were irrelevant to the house purchase: including for example "Hello", "My name is Ginger Smith", "It sure is beautiful outside", and so on, (noise). Three of the statements contained the experimental manipulations being investigated in this research, (hereinafter "the cues"). An example of the procedures, including complete verbal instructions, was completed prior to the actual trials, in order to familiarize the sensors with the procedures. While the sensors were completing this example, a full set of verbal instructions (recorded in the male voice) explained, step by step, what the sensors were asked to do. In addition, the sensors were told that if they were unclear as to the experimental procedures, they should contact the researcher. Contact information was given, however no phone calls for clarification on the experimental procedures were received.

The cues varied on their verbal and paralinguistic signal strength: there were five levels of verbal signal strength: strongly positive (++), moderately positive (+), neutral (0), moderately negative (-), and strongly negative (--). There were three levels of paralinguistic signal strength: positive (+: customer sounds happy), neutral (0: non-expressive tone of voice with little and / or voice inflection), and negative (-: customer sounds unhappy).

While listening to the statements, the sensor would click on one of two buttons: one button was labeled "Information", the other button was labeled "No Information". Sensors were instructed, both in writing and in a recorded statement, to press "Information", if the statement they heard had message content which would help them to estimate customer purchase probability, and to press "No Information", if the statement did not help them in determining that probability. After listening to the complete conversation, sensors were asked to provide another estimate of customer purchase probability: CPPA.

Once the sensors submitted the CPPA information, they were presented with another house description, heard another conversation with three embedded cues, pressed the Information or No Information buttons as they listened and finally provided another estimate of CPPA. Trials were presented in groups of five and in between, demographic information, SOCO or ADAPTS information was collected to prevent sensor fatigue. In total, because this experiment consisted of a 5 x 3 design, the sensors were exposed to 15 different conversations: a full factorial design. House descriptions, noise and cues had been extensively pretested. The order of presentation of the conversations (trials 1-15) was randomized after every 10th sensor and the button location was reversed after approximately every 5th sensor. The order of the presentation of the statements in the conversations was kept constant. Trials started with the statement "Trial will start in three, two, one". Trials ended with the statement "End of trial, please press continue." Finally, sensors completed the SOCO and ADAPTS scales, and provided a variety of demographic information, including a self assessed rating (out of 100), and their income levels,

(scale of 1-5). After completion, the managers of these agents were contacted, and they also assessed the agent on a success scale out of 100. At the end of the test, a page thanked the sensors for their time, and gave contact information. Information on CPP estimates, button clicks, and answers to SOCO, ADAPTS and demographic information were captured in an Access database that was downloaded by the researcher. A flow-chart of the experimental procedures is as follows:



**Figure 2.** Perceptual Chronograph Flowchart

A further testing procedure was identical to the procedures indicated in Figure 2, except rather than clicking on Information or No Information, the sensors had four buttons to choose from: Continue to Listen, Respond to Client, Try to Close, and End Transaction. The button choices were randomized after every twentieth respondent. Button clicks, CPP estimates, and answers were collected in an Access database that was downloaded by the researcher.

### **APPLICATIONS TO OTHER AREAS, LIMITATIONS AND FUTURE RESEARCH**

Although the PC has only been utilized to test the SESER model of communication in the sales context, a number of use extensions for this tool can be postulated. First, as indicated previously, this tool can have applications in the training of salespeople. Expert standards will have been established by the procedures, and novices will have been tested for their responses on the SESER behaviors. The PC data can then be used to compare the standards of the experts with other 'tested sensor', and indicate how the performances vary. Results of this performance could be communicated back to the other tested sensor, either immediately, or in a later feedback session. After this feedback, the other tested sensor could be given more opportunities (as many as necessary) to adjust his responses to approximate the responses of experts. Although it is not likely that there would be a complete re-adjustment in behavioral patterns, it is quite possible that closer approximations to the behaviour of experts can be achieved. This application has not been tested, however, future research will determine the efficacy of this training regime.

Further, as indicated previously, because of the inherent flexibility of the PC, virtually any type of communicated information, including verbal, paralanguage, non-verbal, visual, temporal or contextual information can be included in this media-rich testing tool. The quality of the results, however, will be based on both the sufficiency of the pretesting of the information to be included in the program, as well as the appropriate choice of experimental design. Limitations of the PC include experimental artificiality, and lack of testing in the training context or in other subject fields. However, if the results from the sales context are indicative of potential, the interactive and media rich Perceptual Chronograph appears to be a useful tool in the measurement and testing of perceptual communication responses.

## REFERENCES

1. Ito, T., Larsen, J., Smith, N. & Cacioppo, J. (1998). Negative Information Weighs More Heavily on the Brain: The Negativity Bias in Evaluative Categorizations. *Journal of Personality and Social Psychology*, 75(4), 887-900.
2. Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision Making Under Risk. *Econometrica*, 47, 263-291.
3. Kahneman, D. & Tversky, A. (1995) Conflict Resolution: A Cognitive Perspective, in *Barriers to Conflict Resolution*. In K. Arrow et al, (Eds.) Chapter 2. New York: Norton.
4. Klapper, J. (1960). *The Effects of Mass Communication*. New York: The Free Press.
5. Macmillan, N. & Creelman, C.D. (1991). *Detection Theory: a User's Guide*. New York: Cambridge University Press.
6. Phelps, C. & Hutson, A. (1995). Estimating Diagnostic Test Accuracy Using a Fuzzy Gold Standard. *Medical Decision Making*, 15(1), 44-57.
7. Saber, J. (2004). *Do Hear What I Hear? The SESER Framework of Sales Communication: Listening Skills and Sales Success*. Unpublished doctoral dissertation. University of Alberta. Edmonton, Alberta, Canada.
8. Saxe, R. & Weitz, B. (1982). The SOCO Scale: A Measure of the Customer Orientation of Salespeople. *Journal of Marketing Research*, 19(August), 343-351.
9. Skowronski, J. & Carlston, D. (1989). Negativity and Extremity Biases in Impression Formation: A Review of Explanations. *Psychological Bulletin*, 105(1), 131-142.
10. Spiro, R. & Weitz, B. (1990). Adaptive Selling: Conceptualization, Measurement and Nomological Validity. *Journal of Marketing Research*, 27(February), 61-69.
11. Swets, J. & Pickett, R. (1982). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. New York: Academic Press.
12. Warkentin, M. & Beranek, P. (1999). Training to Improve Virtual Team Communication. *Information Systems Journal*, 9(4), 271-289.
13. Weitz, B., Sujan, H. & Sujan, M. (1986). Knowledge, Motivation and Adaptive Behaviour: A Framework for Improving Selling Effectiveness. *Journal of Marketing*, 50(Oct.), 174-191.
14. Weller, S. & Mann, C. (1997). Assessing Rater Performance Without a 'Gold Standard' Using Consensus Theory. *Medical Decision Making*, 17(1), 71-79.