Toward a Unified Theory of Audiovisual Integration in Speech Perception

Nicholas Altieri

DISSERTATION.COM

Boca Raton

Toward a Unified Theory of Audiovisual Integration in Speech Perception

Dissertation.com
Boca Raton, Florida
USA • 2010

ISBN-10: 1-59942-361-8
ISBN-13: 978-1-59942-361-6

Acknowledgments

I wish to extend my gratitude to the members of my graduate advisory committee, Professors James T. Townsend, David B. Pisoni, John Krushcke, and Jerome Busemeyer. They provided me with the knowledge, theoretical tools, and inspiration that greatly assisted me in carrying out this research. I am also greatly indebted to members of David Pisoni's Speech Research Laboratory and James Townsend's Cognitive Modeling Laboratory for helpful comments and stimulating discussion over the years. The participation of my subjects is also greatly appreciated.

TOWARD A UNIFIED THEORY OF AUDIOVISUAL INTEGRATION IN SPEECH PERCEPTION

Auditory and visual speech recognition unfolds in real time and occurs effortlessly for normal hearing listeners. However, model theoretic descriptions of the systems level cognitive processes responsible for “integrating” auditory and visual speech information are currently lacking, primarily because they rely too heavily on accuracy rather than reaction time predictions. Speech and language researchers have argued about whether audiovisual integration occurs in a parallel or in coactive fashion, and also the extent to which audiovisual occurs in an “efficient” manner. The Double Factorial Paradigm introduced in Section 1 is an experimental paradigm that is equipped to address dynamical processing issues related to architecture (parallel vs. coactive processing) as well as efficiency (capacity). Experiment 1 employed a simple word discrimination task to assess both architecture and capacity in high accuracy settings. Experiments 2 and 3 assessed these same issues using auditory and visual distractors in Divided Attention and Focused Attention tasks respectively. Experiment 4 investigated audiovisual integration efficiency across different auditory signal-to-noise ratios. The results can be summarized follows: Integration typically occurs in parallel with an efficient stopping rule, integration occurs automatically in both focused and divided attention versions of the task, and audiovisual integration is only efficient (in the time domain) when the clarity of the auditory signal is relatively poor—although considerable individual differences were observed. In Section 3, these results were captured within the milieu of parallel linear dynamic processing models with cross channel interactions. Finally, in Section 4, I discussed broader implications for this research, including applications for clinical research and neural-biological models of audiovisual convergence.

TABLE OF CONTENTS

Section 1: Multisensory Integration in Audiovisual Speech Perception

Section 2: Processing Capacity as a Measure of Integration Efficiency

Section 3: Dynamic Modeling Approaches for Audiovisual Integration

Section 4: Conclusions and Future Directions

Section 1

Multisensory Integration in Audiovisual Speech Perception

The information processing approach in the cognitive sciences seeks to understand how the brain processes and then combines segregated external inputs. An area of burgeoning interest in the sensory and cognitive areas of psychology and neuroscience concerns the processing of multisensory stimuli. When distinct inputs enter the system, the output or response must be assessed. The researcher is faced with the task of describing the underlying neuro-cognitive mechanisms that produced the output. The perception of speech serves as one of many interesting examples. Because audition is the dominant recognition modality in normal hearing people, many of us are unaware that speech perception is a multimodal perceptual phenomenon that engages auditory, visual, and even haptic perceptual processes (see Fowler & Dekle, 1991; McGurk & MacDonald, 1976; Sumby & Pollack, 1954).

In a landmark study carried out more than five decades ago, Sumby and Pollack (1954) explored the pivotal role that vision plays in speech perception by showing that lip-reading enhances accuracy scores in spoken word recognition. “Audiovisual enhancement”, as it is sometimes called, is most noticeable when the auditory signal-to-noise ratio is low and speech recognition is difficult in the auditory domain. The authors measured audiovisual enhancement across multiple auditory single-to-noise ratios and observed that the “information transmitted” or visual gain remained relatively constant over a wide range of signal-to-noise ratios.

Another classic perceptual phenomenon in multisensory speech recognition is the McGurk effect—a perception fusion that arises when incongruent auditory and visual signals are presented simultaneously. Perceptual fusions in audiovisual speech integration were first reported by McGurk and MacDonald in 1976 when participants were presented with several types of mismatched auditory and visual signals, including an auditory /ba/ dubbed over a visually articulated [ga]. Participants across age groups typically reported perceiving the fused response /da/ rather than either of the unimodal stimulus categories, /ba/ or /ga/. Several explanations have been advanced to account for perceptual the fusions described by McGurk and MacDonald, including the possibility that visual information related to *place* of articulation is combined with auditory information about *manner* of articulation (see Summerfield, 1987 for a seminal review). Nonetheless, there is currently little consensus in the field regarding the underlying neuro-cognitive mechanisms responsible for producing perceptual fusions and few rigorous biologically plausible theories of audiovisual processing that address dynamical aspects of integration (although see Colonius, Diederich, & Steenken, 2009 for a model of saccadic reaction time).

Experimental findings in behavioral and neuroscience studies continue to be uncovered, yet the theoretical issues and debates that were current decades ago have hardly been settled. The mechanisms that listeners use to extract and combine information from different modalities in real time are not well understood. There have been relatively few studies, new models, or recent rigorous tests of previous models in the area of audiovisual speech perception. The bulk of research uses accuracy as the dependent variable. The approach discussed in this chapter argues that response times (RTs) offer a vital and complementary set of tools that have proven very helpful in determining underlying mechanisms in regions of elementary perception and cognition (e.g., Luce, 1986; Ratcliff & Smith, 2004; Townsend & Ashby, 1983; Van Zandt, 2000; Vickers, 1980; Welford, 1980).

One critical question concerns whether auditory and visual information is processed in separate independent channels before final determination of the linguistic content, or instead is combined early on into a unified code (see Bernstein, 2005; Rosenblum, 2005). This constitutes just one example of a critical question that has been largely unanswered. There is also little consensus in the field regarding the linguistic representations that listeners integrate during perceptual processes. Do listeners, for instance, integrate discrete phonetic features or instead translate the information into some common currency, such as spectral or possibly articulatory/kinematic information (Fowler & Rosenblum, 1991; Summerfield, 1987). The following chapters will be devoted primarily to answering the former question and other related queries.

The issue of whether auditory and visual information are processed in separate independent information bearing channels prior to the final determination of the linguistic content in the auditory and visual channels, or instead is coalesced early on into a unified code is one of the most basic and fundamental questions regarding cognitive operations on speech inputs. It is somewhat surprising that rigorous tests have not been put forth to determine if auditory and visual speech information acquisition actually takes place in parallel instead of a serial manner with distinct processing stages. Intuition tells us that auditory and visual information has to be operated on simultaneously. It is one thing to argue based on intuition and another to prove it or falsify it within the context of a theoretical framework. Cells responsible for sensory transduction and physiological tributaries certainly transmit information in parallel during sensory transduction.[1] However, the functioning of higher order neuro-cognitive mechanisms at the speech recognition level and the real-time spatiotemporal properties remain unknown.

[1] The activation of auditory and visual processing cells occurs in parallel. One possibility is that auditory circuits become active and pass information to visual processing areas and vice versa. Whether processing terminates in one level before information can be passed to the next is an issue that was addressed by McClelland's (1979) Cascade model.

Luckily, there is a way within the domain of reaction times to discover whether bimodal channels are operating in parallel, serial, or are pooled together into a common processing center. It is also possible to examine the stopping rule; whether processing in both the auditory and visual channels must be completed before processing ceases. Workload capacity can also be assessed. This measure can assist in answering the question of whether the level of processing efficiency changes when visual information from the talker's face is provided. Finally, the data can be indirectly informative about cross-channel interactions (Townsend & Nozawa, 1995; Townsend & Wenger, 2004a; 2004b; Wenger & Townsend, 2000; 2001). Few if any determinations of these processing issues have been made in the area of audiovisual speech perception, although some have been made in general studies of multimodal processing using non-speech stimuli (see Berryhill, Kveraga, Webb, & Hughes, 2007; Diederich & Colonius, 2004; Miller, 1982; 1986).

Even current mathematical modeling approaches have not completely clarified the underlying structure and processes inherent in audiovisual speech perception. Modeling efforts are almost entirely limited to explanations and predictions of accuracy. The primary advances in existent accuracy based techniques have been due to a rigorous signal detection approach developed by Braida (1991) known as Pre-Labeling Integration or PRE, and a formidable body of work employing a ratio-of-strengths or evidence formula implemented by Massaro known as the Fuzzy Logical Model of Perception (FLMP) (e.g., 1987; 1998; 2004). These models of audiovisual perception assume that listeners combine information from the auditory and visual modalities and utilize the information in a rather efficient, if not optimal, fashion (Braida, 1991; Grant, 2002; Grant, Tufts, & Greenberg, 2007; Massaro, 2004).

PRE and FLMP are not based on dynamic-in-time processing mechanisms. Taking this significant limitation into consideration, these models clearly cannot adjudicate among positions on major dynamic issues such as parallel versus non parallel processing models. Massaro (2004) at least addressed and acknowledged the issue of parallel "non-convergent" and coactive "convergent" integration in a conceptual manner without proposing a formal test. Interestingly, Braida's Pre-labeling Model of Integration (PRE) assumes that separate sources of auditory and visual information specified by D-dimensional cue vectors are combined during audiovisual speech integration. The audiovisual percept in Braida's model is specified by the cross product of the auditory and visual cue vectors. On some level PRE appears to be construed as a parallel model with separate auditory and visual channels specified in cue vectors. Still, PRE lacks a mathematically specified dynamic framework and the model itself is not subject to rigorous testing on the issue of real time integration and dynamic processing operations.

This research proposes that bringing time-based dependent variables to bear can offer supplementary information to accuracy measures or even information that is unique to the processing of bimodal inputs (see Townsend & Ashby, 1983; Townsend, 1990a; Townsend & Wenger, 2004a for general treatments of these issues). More detail on detection of essential bimodal mechanisms will be given later, but for now I will briefly describe a simplified but critical breakdown of processing models that can provide qualitative and quantitative descriptions of real-time integration processes. This will be followed by a discussion of how these models pertain to theoretical debate about parallel versus coactive processing in the field of speech perception.

Figure 1.1 shows two different "parallel" processing accounts or neural representations of how integration could occur in a bimodal processing system. A schematic account of serial processing will be offered as well. In this case activation cannot begin in the second channel until processing has completed on the first one. The first parallel model, often referred to as a "separate channels" independent processor, assumes that separate decisions or detections are made on each channel. In this framework, the auditory and visual speech streams are processed separately in a simultaneous manner prior to the decision stage. A separate decision is then made on each channel or modality. The latter decision can be formulated using an AND or an OR logic gate. The type of logic gate determines the system's *stopping rule.* Consider the following case. Suppose subjects are given a short list containing numbers or letters and then given a probe item that may or may not be contained in the list. Responding "NO…the probe was not contained in the list" requires one to search exhaustively through each item in the list in order for the system to terminate. In this instance, an AND rule sometimes referred to as a "conjunctive rule" will be in effect and we can say that *exhaustive processing* occurs. Here, the overall RT is determined by the last-terminating channel or item (see Townsend & Ashby, 1983).

Now consider an alternative case. Suppose that listeners are given a task where they have to respond "yes" if presented with the consonant "b" in either the auditory or visual modality. When "b" is presented, each channel accumulates information and when either channel exceeds threshold, the listener responds "yes" regardless of whether the other channel has finished accumulating information. This situation allows the system to cease processing as soon as either feature is identified, and thus leads to an OR decision rule (equivalently, a *disjunctive rule*), which allows a *first-terminating or minimum time stopping time*. In the *coactive* parallel model shown right below the separable channels parallel model, information from each channel is combined into a common information processor that sums activation from each unisensory source. A decision is made once the accrued information in this common processor exceeds threshold.

Lastly, in the serial model described above, processing on the auditory and visual components of “d” cannot occur simultaneously. If the auditory component is processed first then processing in the visual domain cannot begin until processing in the auditory channel has terminated. Once again, if the system has a minimum time stopping rule then a decision can be made when either the auditory or visual channel finishes, whereas if the stopping rule is exhaustive, processing in both channels must finish. The architectures depicted in Figure 1.1 are simple yet elegant in appearing to capture in canonical form some of the critical issues in the audiovisual perception literature. Overall, coactive models represent multimodal processing as proceeding with auditory and visual information before a decision is made on the combined sources of information, whereas ordinary separate decisions parallel processing indicates that decisions on the distinct channels are made. Serial models are of course diametrically opposed to both types of parallel systems and seem a priori unlikely, but nonetheless require falsification in the Popperian sense.

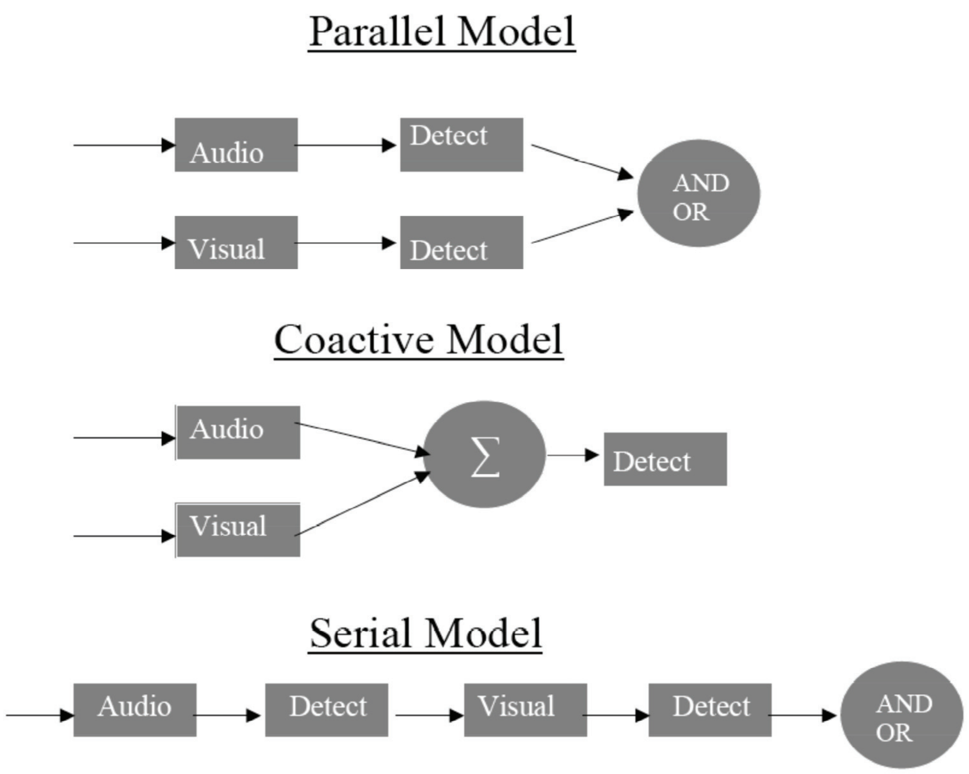

Figure 1.1: Three potential audiovisual processing models. A schematic representation of a parallel model (top) with an OR as well as an AND gate. The coactive model below the parallel model assumes that each channel is pooled into a common processor where evidence is accumulated prior to making a decision. Lastly, the figure shows a serial model that assumes that processing does not begin on the second modality until it finishes processing on the first.

Theoretical Accounts of Audiovisual Integration

I noted earlier that existent models of integration are not temporal in nature and therefore silent with regard to the inherently dynamic processing issues under study here. As we shall see in the following discussion, this section will go forward on the basis that a parallel model with separate decisions is the natural candidate for separate perceptual operations followed by "late" integration. Likewise, coactive models form a natural class of models of "early" audiovisual integration.

Early Versus Late Integration

Rosenblum (2005) argued based on an extensive review of available audiovisual speech data that the neurophysiological and cognitive machinery involved in speech processing extract *amodal* or modality neutral information from the auditory and visual inputs. The theoretical basis of this position is grounded in the gestural theories of speech perception. These theories typically make the assumption that the linguistic representations extracted from the signal are “gestures”. While researchers disagree on many of the finer details, the motor theory of speech recognition (Browman & Goldstein, 1995; Liberman & Mattingly, 1985) and articulatory dynamic theories (Fowler & Rosenblum, 1991; Summerfield, 1987) of speech perception take this basic position. Rosenblum argued that “sensory modality is largely invisible to the speech perception function and the relevant information for phonetic resolution is modality-neutral” (Rosenblum, 2005, p. 51). To that end, he concluded that considerable support for this position comes from evidence showing that the auditory and visual speech streams are integrated in the earliest stages of perception prior to word recognition or phonetic categorization. The main idea here is that each unimodal source of information, both auditory and visual, carries inherent dynamical or “gestural” information at its most basic level. As soon as the speech processing regions of the brain become activated by the input, the underlying dynamical information from each source is extracted and combined naturally since both sources share a “common currency”. The information from the auditory and visual modalities could be combined into a single channel prior to word recognition in which the decision process considers only the sum of the information and not the auditory and visual parameters in the separate modalities. This line of thought appears to be best captured by the coactive model shown in Figure 1.1.

Many behavioral studies have provided a level of support for the theory of “early” integration. As one example, Green and Miller (1985) demonstrated that visually perceived rate of articulation influences the perception of auditory phonetic segments. They showed that visual information about place of articulation affects the perception of voice onset time (VOT). Participants were shown audiovisual clips of a talker saying a syllable that varied auditorially on a continuum from /bi/ – /pi/ (the syllables are pronounced “bee” and “pea” respectively). The corresponding visual information was played either “fast” or “slow”. The results showed that rapidly articulated syllables increased the probability that participants perceived /bi/ rather than /pi/. The authors made sure in a control study that the visual information could not be identified without the accompanying auditory signal. The results suggest that the information from the visual modality affects the speech signal by influencing variable properties of the auditory signal. These results are important because the property that determines whether one perceives the consonant /b/ or /p/ is VOT. Generally speaking, VOT increases the probability that one perceives /b/ versus /p/ or /g/ versus /k/. These results suggest a decision process that has access to both auditory and visual information and combines the two sources of information in the early stages of phonetic perception, although as we shall see, auditory and visual cross channel information sharing can potentially explain these results as well.

Neuroimaging Studies

At this stage of the science, neuroimaging results must be employed cautiously when drawing broad inferences about processing mechanisms that may be widely distributed in the brain. Still, they may offer a potentially useful source of evidence convergent with behavioral findings. At least some neuroimaging evidence from audiovisual speech perception tasks leads to inferences similar to those of the above behavioral results, namely the possibility of coactive processing and early integration (see Calvert & Campbell, 2003).

First, silent lip-reading tasks have been shown to activate left hemispheric regions involved in processing auditory speech stimuli (Calvert, Bullmore, Brammer, Campbell, Iversen, Woodruff, McGuire, Williams, & David, 1997) suggesting that a common code or common underlying information exists across sensory modalities in the speech signal. Calvert and Campbell (2003) also assessed the extent to which visual speech containing dynamic time varying characteristics activates auditory processing regions compared to still-framed images lacking the time-variant components associated with speech. In their study, participants were presented with either sequences of still key frames or moving images of the same duration of a talker saying nonsense syllables. Participants were instructed to look for a visible target syllable such as “voo” in a sequence of other nonsense syllables. Sequences of still key frame images produced activation in the posterior cortical areas associated with the perception of biological motion. Activation was also observed in canonical speech processing areas including Broca’s area and the superior temporal sulcus (STS). However, moving images produced greater activation in these regions compared to still frames. It appears that visual speech accesses areas traditionally believed to be auditory processing regions for language, which is possibly due to neural mechanisms involved in audiovisual integration in the STS (Calvert & Campbell, 2003). Of course, the finding that auditory areas receive information from visual afferents does not imply the latter’s merging with auditory information at that stage, but it is intriguing nonetheless.

Evidence for the existence of neurons that selectively respond to combined audiovisual input also provides evidence that the brain implements an information processing system analogous to the coactive model of integration, where the auditory and visual components of the signal are combined into one channel. One signature for “audiovisual integration” that researchers search for is superadditivity in a multisensory brain region. *Superaditivity* occurs when measured activation in the audiovisual condition is greater than the sum of the activation from the unimodal conditions. The observation of super-additive activation in the superior-temporal-sulcus (STS) indicates the possibility that there are neurons and brain regions responding maximally to audiovisual input (see Calvert, Campbell, & Brammer, 2000). Superadditive activation in multisensory integration areas such as the STS has also been observed in congruent audiovisual speech perception tasks while incongruent audiovisual speech has yielded sub-additive activation (see Calvert et al., 2000). Interestingly, neuroimaging data seems to show that the STS does not respond sub-additively to incongruent or asynchronously presented spoken letters and orthographically presented letters (van Atteveldt, Formisano, Blomert, & Goebel, 2007). These findings suggest that non-speech signals, in this case orthographic targets, and speech sounds might be integrated differently in multisensory brain regions.

The belief that multi-sensory neurons and brain regions are the primary mechanisms responsible for audiovisual integration has not been accepted without question. For one thing, the BOLD response in fMRI designs is a measure of the blood oxygen level in a brain region and represents only an indirect measure of neural activity. fMRI recordings also suffer from poor temporal resolution meaning that it is difficult to ascertain when activation occurs in the time course of language perception. Furthermore, observations of superadditive levels of activation in the STS could be due to the intermingling of unisensory neurons (Bernstein, Auer, & Moore, 2004; Meredith, 2002). This means that areas believed to respond preferentially to audiovisual speech in reality contain large numbers of unisensory neurons that essentially operate in parallel.

This might also suggest that superadditive activation and multisensory neural phenomena might actually arise as a result of statistical effects associated with unisensory neurons (see Allman, Keniston, & Meredith, 2009).

Finally, it is also important to note that the STS is a fairly general multisensory region that responds to a variety of multisensory stimuli such as complex nonspeech gestures (Puce, Allison, Bentin, Gore, & McCarthy, 1998). Puce et al. (1998) observed that when participants were presented with pairs of moving eyes or moving mouths, bilateral activation was observed in the posterior STS. Conversely, the control stimuli consisting of moving checkered patterns failed activate the STS or surrounding areas. Taken together, the upshot of these studies shows that the auditory and visual streams may not be converging to a common processor—or at least that the current evidence in this domain is inconclusive.

The behavioral data are also not unequivocal on the issue of “early” vs. “late” integration. Bernstein (2005) cited several studies showing that audiovisual integration likely occurs in the later stages of processing. One study demonstrated that the introduction of large stimulus onset asynchronies between the auditory and visual modalities fails to abolish the McGurk effect (McGurk & MacDonald, 1976; see also Massaro, Cohen, & Smeele, 1996). This result suggests that a framework assuming extensive unisensory processing can account for audiovisual fusion. There is also considerable evidence showing that while the McGurk effect might appear to be automatic in some individuals it does vary in strength across cultures (Sekiyama, & Tohkura, 1993) and also for familiar versus unfamiliar talkers (Walker, Bruce, & O’Malley, 1995). Bernstein (2005) offered an alternative competing account to the theory information is extracted and combined in the early stages of processing by arguing that neural networks learn predictable associations between auditory and visual information (see Bernstein, Auer, & Moore, 2004). In her framework, extensive unisensory processing occurs in the auditory and visual channels prior to recognition, with audiovisual recognition and information sharing occurring "later" in the perceptual processing stages.

One conclusion from the preceding discussion and review is that the topic of multisensory convergence in speech recognition is far from resolved. I will now turn to description of theory-driven methodology that can adequately assess the theoretical accounts described above as well as other pertinent quantitative issues related to audiovisual speech integration.

The Double Factorial Paradigm: Experimental Methodology for Assessing Models of Audiovisual Speech Integration

Considering the vital distinction between coactive and parallel models of integration in audiovisual speech perception, the next step is to find an appropriate experimental methodology that can distinguish different processing architectures. The double factorial paradigm (DFP) developed over a decade ago by Townsend and Nozawa (1995) is such an experimental tool used just for such a purpose. It can be used to obtain behavioral evidence to distinguish parallel from coactive processing, as well as answering other processing related questions. Descriptions of what we refer to as "coactive" and "parallel" models in the speech perception literature as well as serial mechanisms require more specific mathematical formulation along with behavioral data if they are to be tested rigorously.

One may intuit that if a system ceases processing as soon as a single item is completed the RT signature will be different, regardless of the architecture, from cases where all items must be completed before a response is made. The methodology presented below also assesses the stopping rule; whether every item or channel needs to finish processing before a decision is made or instead whether a final decision can be made when only a subset of the items have been processed. For the sake of brevity, the most efficient rule is always found in the double target data when processing ceases as soon as soon as either target is finished. This is known as the *first-terminating rule*. Hence, the focus on the various architectures endowed with a first-terminating stopping rule. The exception is coactive parallel processing for which the issue is moot. A general treatment of these theoretical issues can also be found in Townsend and Wenger (2004a).

It is also important to stress that the architecture testing methodology does not depend on any specific probability distributions or parameters. The relevant data characteristics are predicted by the various classes of architectures (Sternberg, 1969; Schweickert, 1978; Townsend & Ashby, 1983). An exception is that coactive model predictions rely on Poisson counting models although even these predictions happen to be independent of particular parameter values.

Processing Architecture

The methodology for assessing mental architecture involves systems factorial technology; a methodology where experimental factors are manipulated to assess systems level processing questions. This experimental framework can be used to capture potential interactions between experimental factors. One statistic that has been used to analyze interactions is the mean interaction contrast or $M_{IC} = RT_{ll} – RT_{lh} – (RT_{hl} – RT_{hh})$ (see Sternberg, 1969 for his seminal research assessing serial models). In this formula, "RT" represents mean reaction time and each subscript represents the level of one factor like brightness: h = high, which indicates high saliency and fast reaction times, and l = low, which indicates low saliency and slower reaction times. The hh condition, as an example, might represent auditory and visual stimuli of a high level of clarity that the listener would be able to identify more quickly than if the auditory or visual portions were degraded. An absence of an interaction signifies that the factors are *additive*—a feature that points to serial processing. Subsequent theoretical effort led to extensions of M_{IC} tests to parallel and more complex architectures (e.g., Schweickert, 1978; Townsend & Schweickert, 1989; Schweickert & Townsend, 1989; Townsend & Ashby, 1983). One shortcoming of the M_{IC} that the Double Factorial Paradigm addresses is that the M_{IC} is a coarse measure representing only one point at each level (the mean of the distribution). While an interaction indicates a lack of evidence for serial architecture, parallel processing cannot be inferred based on a non-zero M_{IC} alone.

To address these issues, Townsend and Nozawa (1995; see also Townsend & Wenger, 2004a) developed a more sensitive measure that analyzes the curve of the entire distribution of reaction times referred to as the *survivor interaction contrast* ($S_{IC}(t)$). The $S_{IC}(t)$ is defined as $S_{IC}(t) = S_{ll}(t) – S_{lh}(t) – (S_{hl} (t) – S_{hh}(t))$. Notice that the SIC(t) uses the same sequence of terms as the M_{IC}, only now survivor functions are used rather than mean reaction times. The survivor function S(t), a statistical tool used in *survival analysis*, is a distribution function indicating the probability that a process is still going on. If audiovisual stimuli are presented the S(t) indicates the probability that the word, phoneme, or stimulus has not yet been recognized by time *t*.

The $S_{IC}(t)$ function makes several predictions about processing architecture provided that *selective influence* of experimental factors holds across some interval of time. *Selective influence* refers to the effect that manipulating the level of one factor, such as brightness, has on the processing rate within a channel. Manipulating brightness in one channel should not affect processing in the other channel. When selective influence holds, the reaction times for the high level in a given channel should be faster than reaction times for the low level. The classical method used to test for selective influence is to determine whether the individual survivor functions are “properly” ordered: $S_{ll} > S_{lh}$ (hl) $\geq S_{hh}$. The ordering of survivor functions is a necessary condition for selective influence. Townsend (1990b) proved that an ordering of empirical survivor functions or cumulative distribution functions (CDFs) implies that the means of the distribution are ordered, although an ordering of means does not imply that the survivor functions are ordered. I will now turn to the $S_{IC}(t)$ predictions for specific architectures and stopping rules.

For parallel processing models with independent channels using a first-terminating decision rule the $S_{IC}(t)$ function is entirely positive (Proposition 1 in Townsend & Nozawa (1995)). These models are appropriate to test when completion of any of the processing channels can correctly decide the response. The $S_{IC}(t)$ function for these models is positive since the difference between $S_{ll}(t) – S_{lh}(t)$ should be greater than the difference between $S_{hl}(t) – S_{hh}(t)$. The reason is because the lh trials have an element, namely, the high salience stimulus, that takes less time to process.

A parallel independent model with separate decisions and an exhaustive stopping rule predicts a $S_{IC}(t)$ that is entirely negative (Townsend & Nozawa,1995). This stopping rule is required in cases where all channels must reach completion before it is certain that a correct response can be made. The reason for underadditivity in parallel exhaustive models is because each element must be completed before the system terminates. The processing of the system is determined this time not by the fastest but instead by the slowest element. On the lh or hl trials the longest time tends to be closer to the longest time on the ll trials. The predicted difference between $S_{ll}(t) – S_{lh}(t)$ is therefore smaller than the difference between $S_{hl}(t) – S_{hh}(t)$.

Coactive models with independent channels form a class of parallel models in which the information from each channel is pooled, typically by being added, together into a single channel. While the proofs for $S_{IC}(t)$ functions in coactive models in Townsend and Nozawa (1995) rely on Poisson summation processes, other coactive models such as those based on superimposed diffusion processes, have also been proposed (see Diederich, 1995; Diederich & Colonius, 1991; Miller & Ulrich, 2003; Schwarz, 1994). Simulations of linear dynamic and Poisson models in our lab indicate that the results are general across at least Poisson counter models and Wiener diffusion models (Eidels, Houpt, Altieri, Pei, & Townsend, under revision).