The Categorical Perception of Consonants:
the Interaction of Learning and Processing

Bob McMurray
University of Rochester

Michael Spivey
Cornell University

Introduction

Five decades of research in speech perception and phonetics have been based primarily on a single experimental paradigm. In this paradigm, the participant hears a speech sound (or series of speech sounds) and must report the stimulus as belonging to one of a set of possible response categories. This research has been guided by the basic principle stated in (1):

(1) If the stimuli are varied in the right theoretically motivated ways, the structure of the human speech recognition mechanism will become apparent.

Much research has demonstrated the viability of this basic hypothesis (McQueen, 1996 for examples). The majority of our current understanding of speech perception would have been impossible without this approach and the methodological developments (such as synthesized speech) that it employed.

However, the methodological improvements in fine-grain stimulus generation have not been equally matched by methodological improvements in fine-grain response measures. There has been little development of the standard dependent measure, in which a listener provides a metalinguistic judgment a few seconds after the speech perception event has taken place. A similar state of affairs once existed in the field of sentence processing. Throughout much of the1970's, the dominant method used by researchers to infer the structure of the human sentence processing mechanism was to present a sentence (or series of sentences) and subsequently test the participant's memory for certain aspects (e.g., syntax, semantics) of the sentence. Then, at the 1975 Chicago Linguistic Society meeting, Marslen-Wilson (1975) argued convincingly for developing measures of sentence comprehension that tap representations and processing *during* the comprehension event, rather than after it. This motivation has driven the field of sentence processing for the past two decades, and has resulted in an extremely rich understanding of the possible mechanisms by which a listener/reader integrates linguistic structure and linguistic content in real-time (for recent reviews, see MacDonald, Pearlmutter & Seidenberg, 1994; Tanenhaus & Trueswell, 1995). Moreover, this emphasis on temporal dynamics has spawned attractor network models of sentence processing that integrate learning algorithms

with "settling" algorithms (e.g., Tabor, Juliano & Tanenhaus, 1997), thus allowing the system to simulate the gradual coercion of a temporarily ambiguous signal into a stable-state interpretation.

In the work presented here, we apply that same logic to the study of speech perception. Our goal is to demonstrate that although discrete/symbolic phonetic representations are useful descriptions of the ultimate reportable percept during speech perception, a considerable amount of information is available in the intermediate representations that get computed along the way toward that final state. That is, we do not wish to discount the importance of categorical representations of speech (the existence of categorical speech perception is uncontestable); we merely wish to convince the reader that "getting there is half the fun." Thus, as an addendum to the basic principle (1) above, statement (2) exemplifies this emphasis on the temporal dynamics of speech perception

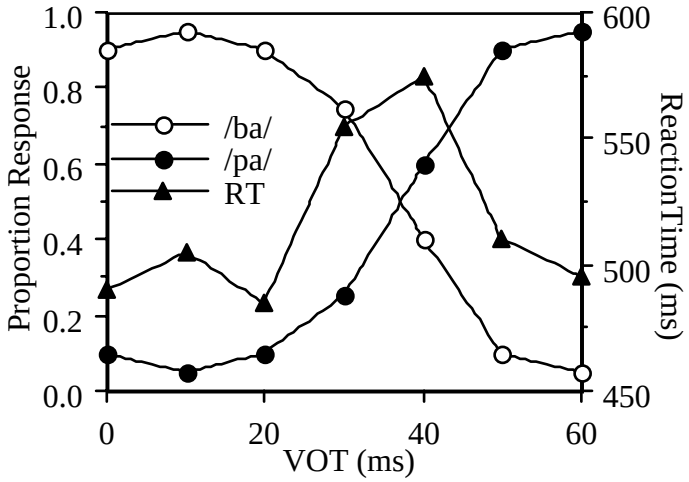
(2) If speech perception is measured *during the perception event*, rather than *after* the perception event, the microstructure of the human speech recognition mechanism will become apparent.

This paper reports on our study of the partial representations that are computed moment-by-moment during speech perception, as revealed by measuring the subject's eye-movements leading up to the selection of a response. In an appropriate visual context, much can be inferred from *where* the subject looks during perception of auditorily-presented linguistic input (e.g., Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995; Allopenna, Magnuson & Tanenhaus, 1998). For example, when a participant is presented a display of several objects, including a candle and a bag of candy, and is then given the spoken instruction "Pick up the candy," they often look briefly at the *candle* before finally fixating the candy and picking it up (Spivey-Knowlton, Tanenhaus, Eberhard & Sedivy, 1998). This "visual cohort effect" suggests a closely time-locked relationship between the real-time accrual of acoustic-phonetic information and the oculomotor mapping of internal representations to objects in the visual environment.

Categorical Speech Perception

The categorical perception of speech sounds is the perfect place to look for graded or probabilistic representations that, in a short time settle into discretely categorized representations because there is strong evidence that the final representations are indeed discrete and, essentially, symbolic (Liberman, Harris, Hoffman & Griffith, 1957). When participants listen to synthesized speech sounds that span the voice onset time (VOT) continuum between /ba/ and /pa/, the lower half of the continuum is consistently identified as /ba/ and the upper half as /pa/. (Additionally, discrimination between different stimuli within a category is at chance.) Thus, the actual VOT of the individual stimulus appears to be discarded, and all that remains in the percept is category membership.

Figure 1: Schematic example of Pisoni and Tash's (1974) results.

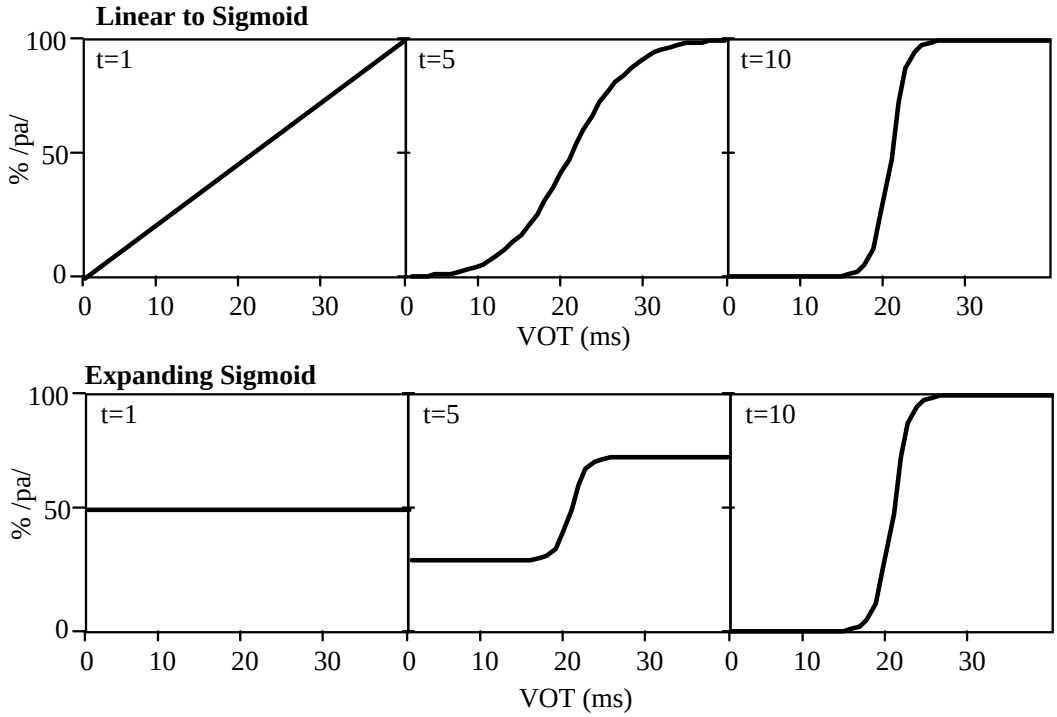


Nonetheless, something very interesting is going on during those few hundred milliseconds between stimulus offset and the reporting of the percept. The first hint at this came from work by Pisoni and Tash (1974), in which they replicated the basic categorical identification of VOT and also recorded participants' reaction times. Stimuli in the ambiguous region of the VOT continuum elicited longer reaction times than those on either half (see Figure 1). Thus, categorization of a particularly ambiguous stimulus takes longer than categorization of a less ambiguous stimulus. This is consistent with the kind of settling, or pattern completion, process seen in attractor networks.

We extended Pisoni and Tash's (1974) exploration of the time course of categorical speech perception by presenting the possible response categories as visual icons on a computer screen that must be clicked with a mouse cursor to make the identification, and recording the participants eye movements during their identification. Since fixations often precede mouse targeting, this would allow us to observe evidence of the speech percept in a state that had not yet "settled" into a discrete category. At this point, we can formulate three hypotheses for the time course of categorical speech perception:

Hypothesis 1: No Change -- The initial state of the speech percept is no different from the end-state, i.e., categorical speech perception is instantaneous and exhibits no temporal dynamics whatsoever.

Figure 2: Two possibilities for the timecourse (left-to-right) of categorical voicing identification. Although the final curve is the same for the two models, the initial and intermediate identification curves are quite different.



Hypothesis 2: Linear to Sigmoid -- The initial state of the speech percept may retain the continuous nature of the synthesized stimuli, as in the top left panel of Figure 2. This continuous perception, indicated by percentage of time that the eyes spend fixating the/pa/ icon, might gradually warp itself over time (t) into a steep sigmoid, as in the top middle and right panels.

Hypothesis 3: Expanding Sigmoid -- The initial state of the speech percept may becompletely ambiguous (see the flat line in the bottom left panel of Figure2), and the halves of the continuum separate themselves over time, retaining their categorical structure (lower middle and right panels of Figure 2).

The eye tracking methodology used here was designed to give us a precise picture of the time course of categorical perception and tell us which of these hypotheses is the best description of the data.

Participants

Subjects were 16 undergraduate students at Cornell University. They were either paid $5.00 or given course credit for their participation. All were native monolingual speakers of American English with normal or corrected to normal vision.

Stimuli

Stimuli were synthesized on a Sun workstation with the Delta system from Eloquent Technologies, Inc. A 9 stimuli /ba/-/pa/ continuum was created by varying the temporal onset of voicing relative to the onset of the release burst (VOT). VOTs varied from -50 to 60 ms with the observed category boundary lying roughly at 10ms. During the post experiment briefing, none of the subjects reported having any trouble identifying the stimuli as /pa/'s or /ba/'s.

Methods

A variant of the visual world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard and Sedivy, 1995) was used to assess the time course of categorical perception. In the visual world paradigm, subjects are asked to perform motor tasks in response to linguistic instructions. In tasks such as reaching or pointing, eye movements have been shown to indicate what the subject is about to reach for or point to. This was adapted to a computer, using mouse control as the motor task.

Subjects were seated at a Macintosh computer and the head-mounted eyetracker (to be described shortly) was calibrated. They were told that they were about to hear a series of synthetic speech sounds and that their task was to categorize them as accurately as possible by clicking with the mouse on one of two large squares labeled /ba/ and /pa/ (which were visible in their correct locations during the instructions). They were told to relax and take their time and that the labeling of the squares would not change throughout the experiment—/ba/ would always be on the left and /pa/ on the right (so that eye movements would not be induced by the subject looking for the button locations).

Throughout the experiment, eye position was monitored with an ISCAN eye tracker. The eye tracker consists of two cameras that are mounted on an adjustable helmet. The "eye camera" records an infrared image of the eye. This image is analyzed by a computer to determine the location of the pupil and the corneal reflection. From this, the computer is able to find the position of the eye relative to the head. This information is combined with the view from the "scene camera" (which records the subject's field of view) as a set of cross hairs indicating the subject's point of fixation. This is computed 30 times per second and recorded on a Sony HI-8 video recorded for analysis.

Each trial starts with the presentation of a single gray circle in the middle of the screen. The subject fixates on this for two seconds to establish that their gaze is midway between the two buttons. When the circle turns red, the subject clicks on it, establishing that the mouse is at the midway point. The circle then disappears, and two gray squares labeled ‘/ba/’ and ‘/pa/’ appear on the computer screen. One of the nine stimuli is played through the Macintosh computer speaker and the subject clicks on the button corresponding to what he or she heard. In addition to eye position, reaction time (the time between the stimulus onset and the mouse response) and the square the subject chose was measured.

Each of the stimuli was presented a total of 7 times, with the order of presentation randomized for each subject. Each trial took approximately 7 seconds, and the experiment as a whole lasted about a half an hour (including time spent setting up and calibrating the eye-tracker). The experiment was designed and run using the PsyScope experimental design software (Cohen, MacWhinney, Flatt & Provost, 1993).

Results

In order to get an accurate picture of the temporal dynamics of categorical perception, eye tracking data was analyzed frame by frame (where one frame equals 33 1/3ms). Research assistants viewed the video tapes of the experiment and for each frame they recorded whether or not the subject was looking at, or saccading to the buttons labeled /pa/ and /ba/ or to a point in space without a button. Saccading to a button was defined as a frame in which eye-gaze was moving in the direction of one of the buttons, but had not reached it yet AND did reach it within two frames. Sound was not recorded on the videotape, so the coders were unaware of which stimulus the subject was hearing.

Averaging this data across subjects and trials for each VOT yielded a picture of the probability of fixation on a particular choice as a function of time. Graphs of three VOTs (-50ms, a good /ba/; 10ms, an ambiguous stimuli; and 60ms a good /pa/) are shown in figures 3, 4 and 5. Qualitiative similarity to eye-tracking results from word recognition (Allopenna, Magnuson & Tanenhaus, 1998, Spivey, 1996) suggest a common underlying process. In particular, the looks to multiple objects immediately after presentation of the spoken stimuli suggest parallel activation of responses, with competition as the disambiguating mechanism.

To verify that the oscillatory eye movements at the beginning of the trial were not the result of random fixation, we used a repeated measures regression. This model examined the effects of VOT (whether it was ambiguous or not) on the number of saccades in the first second of each trial. We limited our data to this small window to ignore any eye movements that were the result of the increased reaction time during the ambiguous trials. The difference between ambiguous and non ambiguous VOTs, though small at 0.127 saccades, was highly significant ($t()=3.175$, $p=.0015$). This correlation between the number of saccades and VOT makes a strong case for the false looks we see in the unambiguous stimuli not being a result of random fixations.

Figure 3: Probability of eye movement as a function of time for VOT=-50

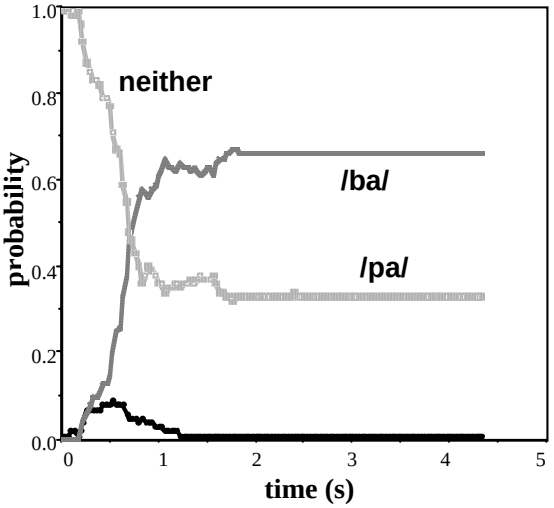


Figure 4: Probability of an eye movement as a function of time for VOT=60

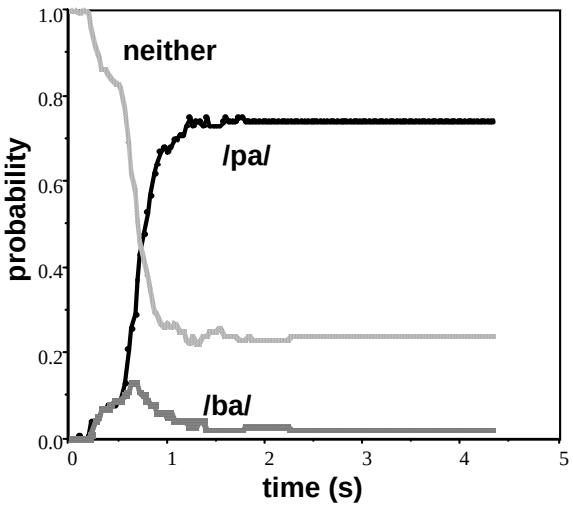


Figure 5: Probability of eye movement as a function of time for VOT = 10

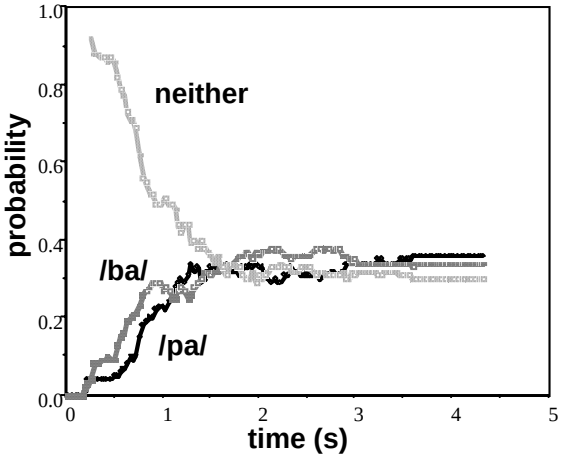
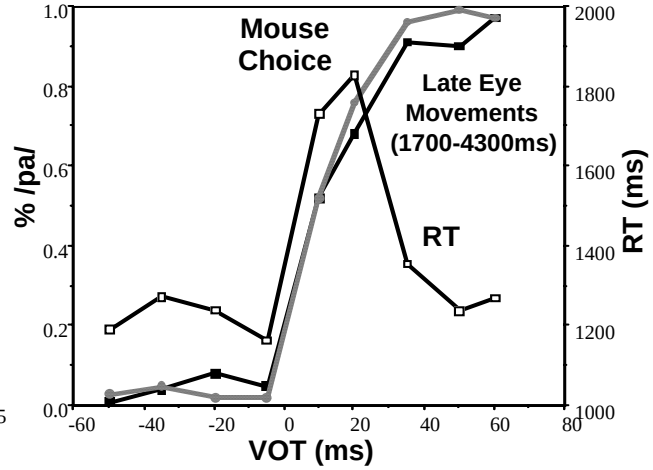


Figure 6: Reaction time, late eye movements and mouse choice as a function of VOT.



Averaging the percentage of looks to /ba/ or /pa/ at several time bins as a function of VOT allows us to view the effect of time on categorical perception phenomena. Figure 6 shows the late eye movements with the mouse choice and reaction time. The pattern of reaction times is qualitatively similar to the pattern found by Pisoni and Tash (1974) and the identification curves exhibit the same steepness. In addition, it is clear that the late eye movements are highly correlated with the mouse choice data. A repeated measures logistic regression (with VOT, trial and whether or not the data came from the late eye movements or the mouse choice) was highly significant significant (yielding a $\chi^2(22)$ of 1538.88., $p<.00001$). More importanlty, no effect was found for whether the data was from a mouse choice or an eye movement ($p > .5$) or for the interactions of this variable with the others (all $p>.4$). This suggests that the two datasets are not different from each other when intrasubject effects and the effect of VOT are partialed out—late eye movements tell us the same thing as the mouse choice.

Figure 7: Percent of stimuli identified as /pa/ as a function of VOT and time.

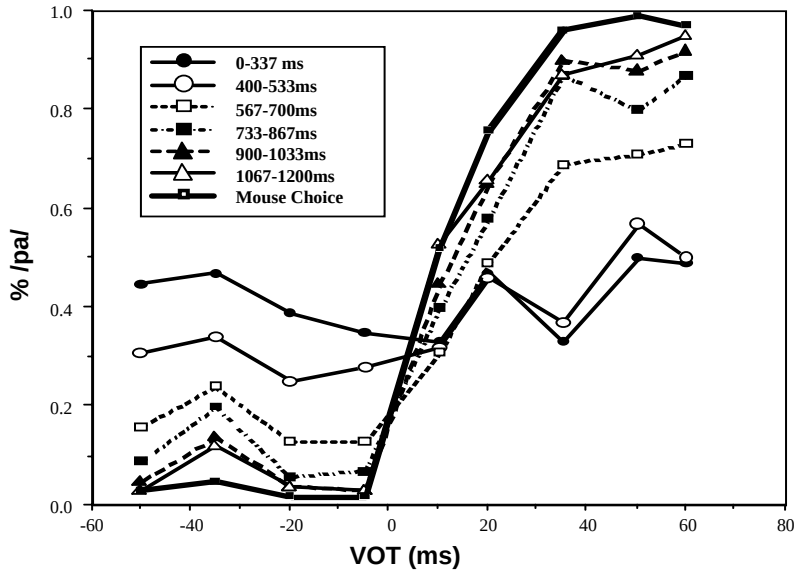


Figure 7 shows identification curves created by averaging the eye movements towards /pa/ at several time-steps. It is clear that category membership is maintained throughout the timecourse. Within categories stimuli are treated "similarly"—i.e. identified with the same probabilities through the process of perception. We'll address these relationships more quantitatively in the next section.

Curve Fitting

The identification function found in categorical perception experiments is closely similar to the logistic function whose formula is given by (3).

$$(3) \qquad \text{logistic}(VOT) = \frac{b_1}{(1 + e^{(-1*(b2*VOT + b3))})} + b_4$$

Roughly, b_1 corresponds to the amplitude of the function, or the difference between the upper and lower asymptotes. The slope or steepness of the function correlates with b_2, and b_3 provides a measure of the location of the category boundary on x axis (VOT). Finally b_4 indicates the y axis location of the lower asymptote, or the height of the function. If each of these parameters is treated as a function of time, we can frame the three hypotheses for how the identification curve might change over time in terms of which parameters are affected by it.

In the no change hypothesis, we would expect none of them to change significantly. The linear to sigmoid hypothesis would expect the slope (b_2) to be small initially and large at the end of the trial (b_3 will have to decrease proportionally as well to maintain the category boundary at the same place). The

expanding sigmoid hypothesis predicts that b_1 would start small and gradually increase (b_4 would have to decrease as well to keep the mean probability at 0.5).

To test these hypotheses we undertook a series of curve-fits. The data were divided into subsets containing the VOT and whether the subject looked at /ba/ or /pa/ at each time step (33ms). So for example, one dataset would include subject #1's eye movement responses at frame 15 (500ms poststimulus) to each VOT. Cases in which the subject looked at neither were ignored in order to capture the relative probabilities of looking at /pa/ as opposed to /ba/). Additionally, eye-movements before 333ms were discarded since the disambiguation point for all stimuli was at 160 ms and it takes at least 150 ms to generate an oculomotor movement.

For each dataset, we mimized the integrated distance between the logistic and the computed probabilities over b1, b2, b3, and b4. Several overlapping constraints were used as well:

1) The value of the logistic at -50 was between 0 and 1
2) The value of the logistic at 60 was between 0 and 1
3) The value of the logistic at -50 was less than the value at 60

Constraints 1 and 2 kept the value of the logistic between 0 and 1 (since the function predicts a probability and cannot therefore be greater than 1 or less than 0). Constraint 3 maintained a positive direction for the function, since there were no cases where the probability of choosing /pa/ was significantly greater than chance given a low VOT. In addition, the following boundaries were used for each parameter:

1)	$0 \le b_1 \le 1$	The amplitude cannot be greater than 1.
2)	$0 \le b_2 \le 10$	The slope could not be negative, nor could it exceed 10 (beyond 10, the actual differences in the curve were very small and caused the curve-fitting algorithm to cycle endlessly).
3)	$-50 \le b_3 \le 60$	The category boundary must lie in the range of the VOTs actually used.
4)	$0 \le b_4 \le 1$	The lower asymptote must be between 0 and 1.

A sequential quadratic programming algorithm was used to fit the logistic curve to each dataset. This yielded values for each parameter for each subject at each time.

Plots of the means of each parameter as a function of time are shown in figure 8. All clearly show a hyperbolic (1/x) relationship to time. Hyperbolic functions of time were fit to each coefficient, yielding the following model:

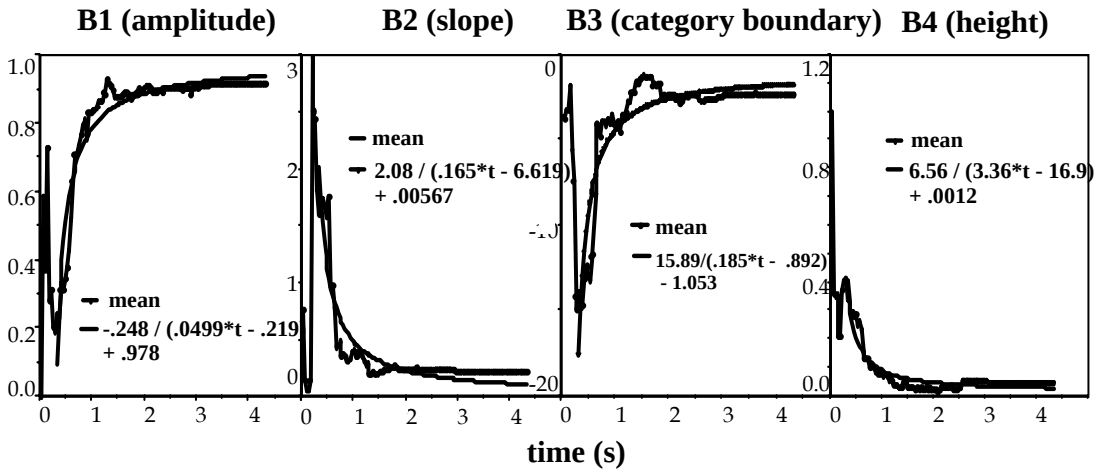
$$(4) \qquad p(\text{looking at /pa/}) = \frac{f_1(\text{time},\beta_1)}{\left(1 + e^{(-1*(f2(\text{time},\ \beta2)*VOT + f3(\text{time}.\ \beta2)))}\right)} + f_4(\text{time},\beta_4)$$

where β_i is a set of four coefficients for the hyperbolic function in 5

$$(5) \qquad f_i(\text{time}, \beta_i) = \frac{\beta_{i1}}{(\beta_{i2} * \text{time} + \beta_{i3})} + \beta_{i4}$$

To test the model's fit, a logit model was used with predicted probability as the dependent variable and actual looks as the independent variable. A \technique analogous to hierarchical regression was used to assess the role of each of the coefficients. When the four logistic coefficients were held constant at their means (the logistic curve did not vary as a function of time) the model yielded an Nagelkerke estimated R^2 of .659 (a standard R^2 cannot be used with a logistic regression) which was highly significant ($\chi^2(4)=52844$, $p<.00001$). Next, amplitude and height (b_2 and b_3) were allowed to vary as a function o f time. This yielded an R^2 of .67 (also significant at $p<.00001$). Importantly, a χ^2_{change} statistic was significant as well ($\chi^2_{change}(6)=1305$, $p<.00001$)—allowing time to affect amplitude did account for significant new variance. In the last step, slope and category boundary were added to the model (b_1 and b_4). This yielded a total R^2 for the model of .672. Although this change was significant ($\chi^2_{change}(6)=221$,

Figure 8: Plots of the mean values and fitted hyperbolic functions of each of the four parameters of the logistic function as a function of time.



$p<.00001$), the amount of additional variance was much smaller than when amplitude was entered into the equation (0.2% as opposed to 1.1%), suggesting that slope played a very small role in the model. Additionally, when slope was allowed to change as a function of time *without* amplitude, the model actually performed worse than when all four coefficients were held constant ($R^2_{slope}=.655$ vs. $R^2_{none}=.659$). This suggests an interaction between amplitude and slope change such that changing slope as a function of time only improves this fit if

amplitude is changed with it. Unfortunately, this statistical technique could not adequately be explored with these techniques.

It seems clear from these fits that our third hypothesis is true. The amplitude of the logistic function is clearly increasing, and although slope is affected by time, it is not in the direction specified by hypothesis 2—slope changes from steep to shallow. Additionally it does account for much variance taken with amplitude, and by itself, does worse than the model, which does not include time. This technique has not only allowed us to determine which of our three descriptive hypotheses is correct but also given us a very precise picture of the time course of categorical perception. They have shown us that most of the variance can be accounted for by hyperbolic change in the amplitude of the logistic as a function of time, and that although there is a slight change in slope (again hyperbolic), it is from steep to shallow, not the other way around.

Models

Now that we understand fully what is occurring during the time course of categorical perception, it is important to begin to address the psychological consequences of these findings. To do this we instantiated each of our three hypotheses in a connectionist network. Since we know the architectures of each network and the kind of representations and processing they use, if we can find one that fits, we can make a case for the psychological processes that might be responsible for temporal dynamics of categorical perception.

The "no-change" hypothesis was instantiated in a classic perceptron. This network is a two layer feed-forward network that learns statistical distributions of its input using competitive Hebbian learning (Rumelhart and Zipser, 1986). Forty input and output nodes were used with the input array indicating VOT—lower indexed nodes representing small VOTs and higher nodes large VOTs. Input was in the form of a spike chosen from a bimodal normal distribution (as per Lisker and Abramson, 1964). The output was "winner take all"—after the output was computed, the node with the highest activation was given all the activation and the rest of them set to zero. Learning occurred after this idealized "competition" using a variant of Runelhart and Zipser's (1986) unsupervised learning rule and based on the ideas of Hebb (1949).

(6) $\Delta W_{io} = (I_i * O_o - W_{io}) * \varepsilon$

ΔW_{io} refers to the change in weight connecting input node i (I_I) and output node o (O_o). W_{io} refers to the current connection strength and ε is the learning rate (set to .1 for this simulation).

After 5000 training epochs, the network correctly learned to categorize VOTs into one of two categories. This categorization was in the form of one of the 40 output nodes representing voiceless sounds and one indicating voiced sounds. Although it may seem that only two nodes were needed, previous research (McMurray, in preparation) has suggested that additional output nodes

Figure 9: Identification function and processing time across VOTs

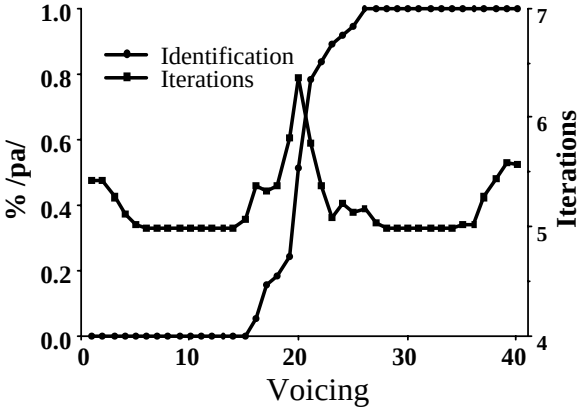
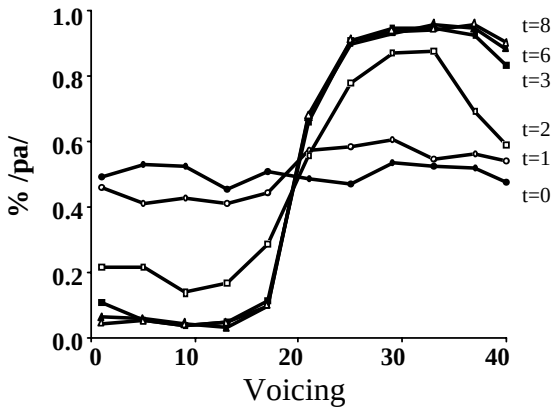


Figure 10: Percent of stimuli identified as /pa/ as a function of time and VOT.



are necessary for the learning process. Although this model did show the correct categorization behavior, it showed none of the temporal dynamics we were interested in. In short, this network was extremely sensitive to the statistics in its learning environment but had no temporal processing.

To model the linear to sigmoid hypothesis, the Normalized Recurrence Network was used (McRae, Spivey-Knowlton & Tanenhaus, 1998, and Spivey & Tanenhaus 1998). This network consists of two input nodes and two output nodes. Input nodes indicated the probability that the VOT is a /pa/ or a /ba/ (with the decision indicated by the output nodes). A very good /pa/ for example might have 0.9 and 0.1 as the activation for it's input nodes, where a /ba/ would be 0.1 and 0.9 and an ambiguous stimuli would have 0.5 and 0.5. At each round, the sum of activation at either level can only be one, and the network cycles through a competition algorithm until the change in output nodes is less than 0.0001. This network does categorize correctly (since the categories are built into the architecture) and shows the appropriate increase in reaction time at the category boundary (since it take more iterations for the competition algorithm to resolve an ambiguous input). However, since the network is not sensitive to the statistical distributions of its input, it does not treat inputs from the same category the same—perceptions start out as a linear function of VOT and are categorized over time (as per the linear to sigmoid hypothesis). This network showed no statistical sensitivity, but did show competitive processing.

Our third hypothesis, the expanding sigmoid was instantiated in a new type of network, the Hebbian Normalized Recurrence Network. This network combines the representation and learning of the perceptron with the processing algorithm of the normalized recurrence network. Like the perceptron, the Hebbian normalized recurrence network has 40 inputs and outputs. Again, the input is chosen in the same manner as the perceptron (from a bimodal normal distribution). After activating the input nodes, the network then uses the following algorithm:

1) The inputs are normalized so that they sum to 1.
2) The outputs are computed by multiplying the inputs by the weight matrix and adding that value to the current value.
3) The outputs are normalized so that they sum to 1.
4) The weights are modified using the Hebbian Learning Rule.
5) The output is multiplied by the weight matrix transposed and this value is multiplied by the input vector and added to it.
6) This repeats until the average change in the outputs is less than .00001.

This algorithm is similar in nature to that of the normalized recurrence network, except that the addition of the weight matrix allows for a topographic representation of the input, as well as the possibility for learning. Where the normalized recurrence network can best be thought of as a system of two equally sized and spaced attractor basins, the Hebbian normalized recurrence network learns those attractor basins and can learn as many as it needs in whatever "shapes" it needs. It shows sensitivity to both the statistical distribution of VOT in the learning environment and competitive processing.

The output of the Hebbian normalized recurrence network is shown in figures 9 and 10. It is clear that it learns to categorize the input correctly and shows the appropriate spike in processing cycles at the ambiguous stimuli. In addition, it also seems to have the properties of the expanding sigmoid in that category members are treated similarly throughout the time-course of processing. In fact the only deviation from the patterns of data we saw empirically is the tendency of the model to show graded category membership at the far ends of the VOT continuum early in the time course (at the extreme of devoicing and prevoicing). This is a testable hypothesis that we have begun to look at by exploring the extreme devoiced and prevoiced ends of the VOT continuum.

This network also shows a high degree of neurological plausibility. The learning algorithm has been shown to have a close correlate with long term potentiation a process by which synaptic connection strengths increase after concurrent firing by pre- and post-synaptic neurons. In addition, the competition algorithm is based on Heeger's work (1993) in neural interaction, and so is also supported neurologically. For these reasons, this "breed" of connectionist models may be an excellent way to explore the effects of both learning and processing.

Conclusions

Through the bidirectional interplay of experimentation and simulation, this work directs our attention to an aspect of categorical speech perception that has been all but ignored: the temporal dynamics of perception. From the eye movement data, it appears that the continuous information of VOT for a given speech stimulus may indeed be discarded rather quickly during the perception process. However, complete categorization is not instantaneous. Early on in speech perception, each half of the VOT continuum is treated as belonging *somewhat more* to one category than the other. As time proceeds, this gradual categorization becomes

more confident and discrete, until finally displaying the signature step-function of categorical speech perception.

This pattern of data could not be simulated with a neural network that simply incorporated a statistical learning algorithm but no temporal dynamics. Similarly, the results could not be simulated by a hand-coded attractor network that displayed temporal dynamics but no statistical sensitivity. A neural network that combined competitive Hebbian learning (Rumelhart and Zipser, 1986) with a "settling" algorithm (Spivey & Tanenhaus, 1998) provided the only satisfactory account of these data. This network also opens the door to further explorations, both in the extreme ranges of VOT, as well as in issues regarding the development of speech perception (McMurray, in preparation).

These findings and their accompanying simulations have broad implications for speech processing and phonetics in general. If it actually takes a few hundred milliseconds to discretize one's percept of a potentially noisy consonant, speech recognition is an even more convoluted process than initially suspected. Before one phoneme is fully categorized, the next few are already being received as input. Mapping such a string of multiple partially active and mutually exclusive phonemic representations onto possible lexical items will no doubt be a massively parallel process. This perspective lends support to feature-based parallel-activation models of speech recognition (e.g., McClelland & Elman, 1986), in which phonetic features are not binary but exhibit graded activation levels. Thus, treating phonetic representations as discrete logical symbols may be useful for idealized instances where noise and interfering signals are absent. However, when speech perception is considered in realistic noisy environments, with the real-time accrual of acoustic-phonetic input being faster than the real-time classification of that input, phonetic representations will have to be treated as probabilistic representations.

References

Allopenna, P., Magnuson, J., and Tanenhaus, M. (1998) Tracking the time course of spoken word recognition using eye-movements: evidence for continuous mapping models. *Journal of Memory and Language*, 38(4) 419-439.

Cohen J.D., MacWhinney B., Flatt M., and Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25(2), 257-271.

Heeger, D. J. (1993) Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology*. 70(5), *1885-1898*

Hebb, D. (1949) *The Organization of Behavior*. New York: Wiley

Liberman, A.M., Harris, K.S., Hoffman, H.S., and Griffith, B.C. (1957) The Discrimination of Speech Sounds Within and Across Phoneme Boundaries. *Journal of Experimental Psychology* 54(5), *358-368*

Lisker, L., and Abramson, A. (1964) A Cross Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word*, 20 *384-422*

Marslen-Wilson, W. (1975) The limited compatibility of linguistic and perceptual explanations. In *Papers from the parasession on functionalism*. Chicago Linguistic Society.

MacDonald, M., Pearlmutter, N. and Siedenberg, M. (1994) The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101(4), *676-703*.

McClelland, J. and Elman, J. (1986) The TRACE Model of Speech Perception. *Cognitive Psychology*, 18, *1-86*.

McMurray, B. (in preparation) The Hebbian Acoustic/Phonetic Category Acquisition Model.

McQueen, J. (1996) Phonetic Categorization. *Language and Cognitive Processes*, 11(6), 655-664

McRae, K., Spivey-Knowlton, M. and Tanenhaus M. (1998). Modeling the effects of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 37, *283-312*.

Pisoni, D., and Tash, J. (1974) Reaction times to comparisons with and across phonetic categories. *Perception and Psychophysics* 15(2), 285-290

Rumelhart D., and Zipser, D. (1986) Feature discovery by competitive learning. In Rumelhart, D., McClelland, J., (eds). *Parallel Distributed Processing: Exploration in the Microstructure of Cognition. Vol 1*. Cambridge, MA: the MIT Press *151-193.*

Spivey, M. (1996) *Integration of Visual and Linguistic Inforamtion: Human Data and Model Simulations*. PhD Dissertaion, University of Rochester

Spivey, M. and Tanenhaus, M. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, *1521-1543*.

Tabor, W., Juliano, C., Tanenhaus, M. (1997) Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing, *Language & Cognitive Processes* 12, *211-271*.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., and Sedivy, J. (1995) Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), p 1632-1634

Tanenhaus, M., Trueswell, J. (1995) Sentence comprehension. in Miller, Joanne L., Eimas, Peter D. (eds) *Speech, language, and communication. Handbook of perception and cognition (2nd ed.)*, San Diego, CA: Academic Press, *217-262*.

Acknowledgements

The authors would like to thank Tobey Doeleman for help with the speech synthesis, Michelle Spence, Rebecca Mabie and Melinda Tyler for coding the data, Harry Reis for help with the curve fitting, Richard Aslin for helpful comments, and Robbie Jacobs and JC Ducom for computer time.