31 Speech Production and Perception in Patients with Cerebellar Lesions

Richard B. Ivry and H. S. Gopal

Over the past few years, we have sought to identify the elementary mental operations that underlie the performance of coordinated behavior. The putative operations studied here include ones related to the control of sequencing, force, and timing. A guiding hypothesis of our work is that success or failure in the execution of movements may, in part, reflect individual differences in these computations. Moreover, we have sought to determine whether these operations are independent of one another, or whether they should be construed as the unified computational focus of the motor control system (Keele and Ivry 1988).

Elucidating elementary mental operations in the study of motor control entails answering two central questions. Of primary concern is what the operations are and how these operations interact. In addition, a natural extension of this concern is to consider the generality of a hypothetical operation. For example, if the evidence suggests that timing is controlled independently from force in reaching movements, a logical question to ask is whether the same timing mechanism is used in other effector domains.

Figure 31.1 presents two ways to conceive of how elementary operations might be organized. They deal with a case in which the behavioral goal is invariant; an actor is trying to attract the attention of another person. Figure 31.1a sketches how a simplified model of the motor system might divide its functions in terms of effector systems. According to this view, many elementary computations must be performed for an action to occur, but the computations are specific to effector systems, perhaps because of their idiosyncratic constraints. Figure 31.1b retains the notion that there are elementary operations, but here the operations are shared across effector systems. Not only is a specialized module activated when a behavioral goal requires this module's particular computation, but the same module can be invoked by other effector systems when that computation is needed. In Fodor's (1983) terms, figure 31.1a embodies a system characterized by both vertical and horizontal modularity. Vertical modularity reflects the segregation of function across task domains, and horizontal modularity reflects the fact that within each domain there are distinct computational modules. Vertical modularity is violated in figure 31.1b because the same computational modules are shared across domains.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

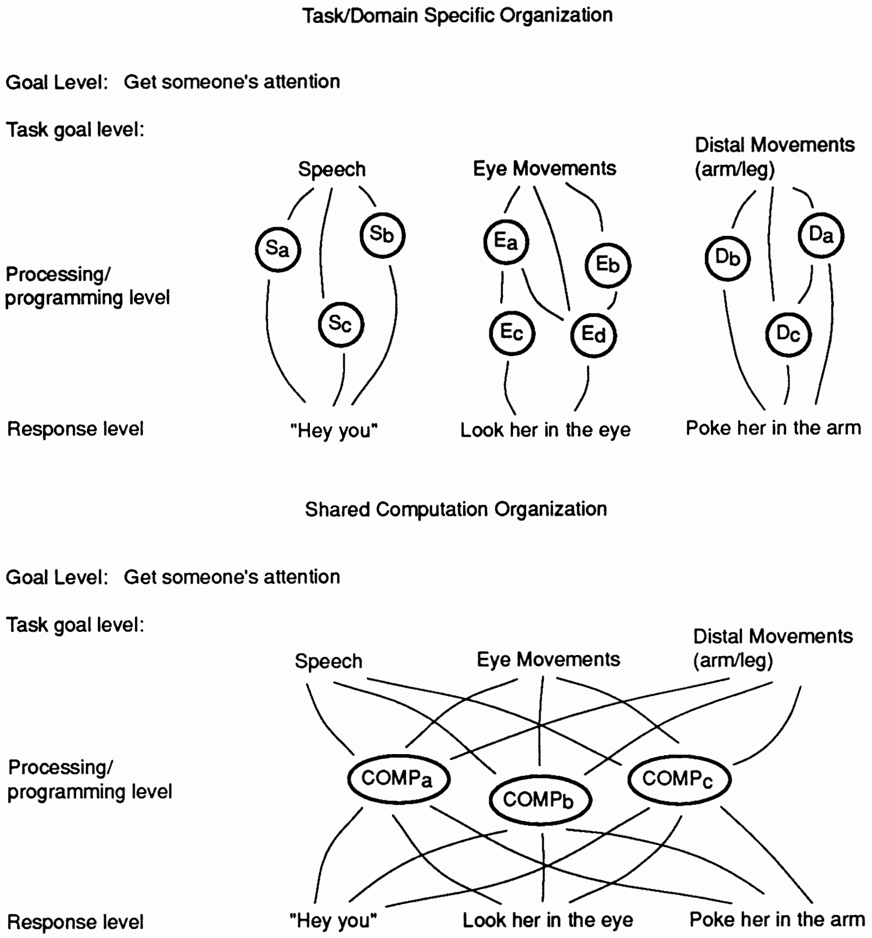


Figure 31.1 Two hypothetical organizations of the motor system. Figure 31.1a entails vertical and horizontal modularity. Figure 31.1b entails only horizontal modularity.

Much past research in motor control has either ignored the notion of separable operations or assumed, perhaps for practical reasons, the organization depicted in figure 31.1a. For example, many researchers have sought to identify the minimal set of parameters needed to describe a movement trajectory (e.g., Flash and Hogan 1985; Hasan 1986; Lacquaniti and Soechting 1982). Generally, the computations of these parameters require joint information about forces and time, and they also frequently require knowledge about the specific effector system (Atkeson 1989). In addition, researchers have generally studied performance of specific tasks and, thus, have limited their attention to single effector systems.

By contrast, our research strategy has been to test directly the viability of the organization suggested in figure 31.1b. We have used a number of different methodologies to delineate a set of candidate elementary operations and to assess their generality. In correlational studies, we have found evidence that timing and force control might constitute two independent operations.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

For example, Keele, Ivry, and Pokorny (1987) had subjects perform two tasks. The first task required subjects to produce a series of isochronous intervals by repetitive tapping. The second task assessed the subjects' ability to produce consistent force pulses. The movements were performed with either the finger or the forearm. Correlations were low between performance with the same effector for timing versus force production. However, the correlations were high within tasks. For example, a subject good at controlling force in finger flexions was also typically good at controlling force in forearm extensions. The same was observed for the timing task. In addition, we have found that timing precision on motoric tasks also correlates significantly with perceptual timing acuity (Keele et al. 1985).

Research with neurological patients has further supported this dissociation between timing and force control. In particular, we have found that patients with lesions of the cerebellum have difficulty producing a series of isochronous intervals (Ivry and Keele 1989). Moreover, an analysis of patients with focal lesions indicated that when the lesions include the neocerebellum, this difficulty stems from increased variability in a central timing process rather than from a deficit in implementing a central movement command (Ivry, Keele, and Diener 1988). On the other hand, patients with lesions of the basal ganglia are unimpaired in time production, but have difficulty in regulating the force of a movement (Hallett and Khoshbin 1980; Ivry 1990; Stelmach, Teasdale, Phillips, and Worringham 1989). Together, these studies indicate that the basal ganglia and cerebellum are involved in separable computations for the production of coordinated behavior, thus supporting modularity as depicted at the processing/programming level of figure 31.1 (also see Lundy-Ekman et al. 1991).

Additional studies lead us to believe that horizontal modularity (fig. 31.1b) provides a more parsimonious functional description of the motor system than does vertical modularity (fig. 31.1a). Of greatest interest and surprise is the finding that cerebellar lesions not only disrupt the production of periodic movements but also impair subjects' ability to estimate the duration of intervals demarcated by auditory clicks (Ivry and Keele 1989). This provides the first demonstration of a role played by the cerebellum in a non-motoric task. The duration-perception deficit is not generic; patients who show it are unimpaired in judging the loudness of similar stimuli. Thus, we have argued that one function of the cerebellum is to operate as an internal clock for both perception and production (Keele and Ivry 1991; cf. Gordon and Meyer 1984).

Given these results, we have recently examined the performance of cerebellar patients on a new perceptual task, velocity perception, which may require precise timing (Ivry and Diener in 1991). Here, each stimulus was composed of a line of dots that continuously moved across the field of view. The subjects judged whether the velocity of a test stimulus was faster or slower than the velocity of a standard stimulus. In three experiments sampling a range of velocities, patients with cerebellar lesions were impaired in making the velocity judgments. Control experiments indicated that the visual deficit was specific to velocity judgments and could not be explained by faulty eye

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

movements. These results provide a second demonstration of the timing functions of the cerebellum in perception. The utilization of this computational capability in tasks as diverse as finger tapping, auditory time perception, and velocity perception corresponds to a violation of vertical modularity (Fodor 1983).

To further explore the generality of the cerebellar timing system, the present chapter reports new experiments that extend our cerebellar research into the linguistic domain. This extension is a natural outgrowth of earlier work. Temporal regularities are pervasive in speech perception and production. These phenomena have been studied at the level of phonemes (House 1961; Lisker and Abramson 1967; Miller and Liberman 1979), syllables (Kent and Moll 1975; Kozhenvnikov and Chistovich 1965), and larger segments of speech where rate and stress parameters vary (Kelso et al. 1985; Tuller and Kelso 1990; Weismer and Fennell 1985). The question of whether timing in speech is explicitly represented or is an indirect consequence of biomechanical and dynamical constraints has proven difficult to answer (see Fowler 1980; Tuller, Kelso, and Harris 1983).

Examining the performance of cerebellar patients with deficits in timing should prove useful in answering a number of questions related to this one. First, can the speech disorders observed in these patients be described by the hypothesis that the patients have suffered a selective deficit in temporally coordinating the numerous neuromuscular events involved in *producing* speech? Second, are the timing capabilities of the cerebellum involved in speech *perception*?

Evidence of a perceptual deficit would support the hypothesis that some psychological processes directly use the temporal information in speech and that these processes invoke the timing capabilities of the cerebellum. On the other hand, the psychological processes involved in speech perception may not directly exploit temporal cues identified through the acoustic analysis of speech. Temporal cues in the acoustic signal may be correlated with spectral cues (e.g., Port 1981) or other sources of information such as those defined at the level of the motoric actions that produce the sound (Abbs and Gracco 1983).

31.1 EXPERIMENT 1

Experiment 1 investigated the performance of cerebellar patients on two speech perception tasks. The stimuli for one task formed a temporal continuum; for the other task, the stimuli formed a spectral continuum. If the cerebellar timing system is required for processing temporal information in speech, we expected the patients to be impaired in judging the temporal continuum. The spectral continuum was included for control purposes.

The temporal continuum was established by varying the voice-onset time of stop consonants. All stop consonants in syllable initial positions are initiated by the rapid release of airflow from a point of constriction in the vocal tract.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

One dimesion by which these consonants can be classified is voicing. For voiced stop consonants such as */b/*, */d/*, and */g/*, the onset of vocal cord vibration (voicing) occurs at approximately the same time as the release of airflow. In contrast, voicing is delayed for an interval of time after the release of consonantal airflow for voiceless stops such as */p/*, */t/*, and */k/*. The temporal difference between the release of airflow and the onset of voicing is the voice-onset time (VOT). For example, in English, VOT is generally less than 20 ms for voiced stop consonants and greater than 40 ms for voiceless stop consonants. Although many languages make a voiced/voiceless distinction, the mean values differ across languages and individuals (Lisker and Abramson 1964). Moreover, some languages contain consonants that are in a different region of the voicing continuum, namely, prevoiced such that vocal-cord vibration precedes the release of air at the oral tract (Lisker and Abramson 1964; Strange and Jenkins 1978).

Discrimination between voiced, voiceless, and prevoiced consonants has been examined in studies of categorical perception (see Studdert-Kennedy 1976 for a review). For example, Abramson and Lisker (1968) generated three series of synthetic consonants that varied in VOT. Within each series, the frequencies were constant, simulating formants associated with one of three places of articulation: labial, alveolar, and velar consonants. Across all three continua, listeners produced identification and discrimination functions indicative of categorical perception. While the mean boundary was around 30 ms, the exact boundaries varied slightly as a function of place of articulation, Nonetheless, these results support the hypothesis that voicing is a salient feature in speech perception.

In the current experiment, patients with cerebellar lesions were tested on a series of consonants that differ in VOT. The cerebellum has proven critical for accurate timing in perceptual and motoric tasks. Thus, our primary interest was whether these patients would show evidence of categorical perception. If the cerebellar timing mechanism is involved in the computation of VOT, then damage to it should reduce the categorical effect. On the other hand, null results would suggest one of two interpretations. The psychological mechanism underlying categorical perception of the voicing dimension may not involve a temporal computation, or this computation may be outside the domain of the cerebellar timing system.

As in all patient research, we must be cautious in making claims about specific functional capabilities from the study of pathology. Patients are notorious for showing aberrant performance on a wide range of tasks, so an experimenter must provide a means to determine the specificity of a deficit. For example, an observed deficit may reflect impairment in a timing mechanism. However, it could also reflect a deficit in auditory processing or a general inability to follow directions for a psychophysical test.

To control for this possibility, the patients were also tested with a continuum of consonants that varied in place of articulation, ranging from */ba/* to */da/*. Acoustically, this continuum is primarily formed by varying differences

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

in the frequencies of the second and third formants during the transitional portion of the consonant. VOT is held constant. Thus, unlike the *ba-pa* series, there is no obvious temporal cue distinguishing the two categories of the *ba-da* series. Therefore, we expect cerebellar patients to be unimpaired in classifying these syllables if the deficit in VOT discrimination is a specific one. It should be kept in mind, though, that formant transitions convey both temporal and spectral information (e.g., Miller and Liberman 1979). So, the *ba-da* series provides only a partial control test. If the patients were to perform poorly on both continua, additional control tests would be needed.

Method

Subjects All subjects were native speakers of German. The experiment was conducted at the Neurology Clinic of the Tuebingen University Medical School in Tuebingen, Germany. Fifteen patients with cerebellar disorders were tested. Many of them have been described in previous reports (Ivry and Keele 1989).

Two of the patients had focal lesions from stroke, and one had a bilateral ischemic lesion. The remaining patients had symptoms consistent with a diagnosis of cerebellar atrophy ($n = 8$) or olivopontocerebellar atrophy ($n = 4$). The difference between these latter two diagnoses is more a matter of degree than a reflection of distinct entities. In all of the cases, the diagnosis was consistent with neuroimaging results obtained with CT or MRI scans. The patients ranged in age from 22 to 75 years (mean = 51.6, SD = 16.7).

The patients received a clinical examination by a neurologist at the time of testing. The examination included tests of postural stability, oculomotor reflexes, voluntary movements with the upper and lower extremities, and speech. Though the patients varied in terms of their performance on these tests, all showed deficits on some part of the examination. Indeed, most of the patients had deficits in tests of postural and voluntary motor control.

Seven other people who had no history of disease of the central nervous system were recruited as control subjects. These subjects were clerical workers at the hospital ($n = 5$) or patients with back injuries ($n = 2$). They ranged in age from 24 to 56 years (mean = 44.4, SD = 13.0). Our only criteria in selecting the control subjects were that they were native German speakers and that the group spanned the approximate age range of the patients.

Stimuli The stimuli were generated on a PDP 11/73 computer with the cascade-parallel synthesizer designed by Klatt (1980) and modified by Kewley-Port (1978). The synthesizer was set to cascade mode.

Each syllable had a total duration of 250 ms with identical frequencies during its steady-state portion. The steady-state formant frequencies were: F1 = 750 hz, F2 = 1200 hz, F3 = 2350 hz, F4 = 3300 hz, F5 = 3850 hz, and F6 = 4900 hz. The fundamental frequency for each syllable began at 121 Hz, rose to 125 Hz by 40 ms, and then fell linearly to 100 Hz at offset. All of the stimuli were made without release bursts.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

There were two sets of nine stimuli each. One set formed the *ba-pa* continuum, and the other formed the *ba-da* continuum. Figure 31.2 shows examples of the stimuli, depicting the fundamental frequency and the first three formants. The *ba-pa* continuum was created by varying the energy onset for F0 and F1 relative to the onset of F2 and F3. For the end point /*ba*/, the onset of F0 and F1 preceded the onset at the higher frequencies by 10 ms. The onset of F0 and F1 was delayed 70 ms for the end point /*pa*/. The rest of the series was synthesized by varying the onset of F0 and F1 in steps of 10 ms, thus creating a continuum of VOT values ranging from −10 ms to +70 ms. The starting frequencies for F1, F2, and F3 were 200 Hz, 825 Hz, and 2000 Hz, respectively. The formants followed linear trajectories, achieving their steady-state values by 35 ms. There were no transitions for the higher formants.

For the *ba-da* series, the energy onset of all the formants occurred simultaneously. The onset frequency values of F2 and F3 were manipulated, while all other aspects of the syllable were held constant. The end point /*ba*/ was the same as the stimulus with VOT = 0 in the *ba-pa* series. The initial frequencies of F2 and F3 were 825 hz and 2000 hz. The initial frequencies of the end point /*da*/ were 1500 hz and 2630 hz. Thus, F2 and F3 had rising trajectories for the endpoint /*ba*/ and falling trajectories for the end point /*da*/. The series was created by varying F2 in steps of 84 hz, and F3 in steps of 79 hz.

Four blocks of stimuli were made for each series. Each block consisted of eight repetitions of the nine stimuli, yielding a total of 72 syllables. Stimulus order within a block was random. A four-second interstimulus interval separated the syllables. In addition, two training blocks were created with only the end point stimuli. These blocks contained ten repetitions of each end point.

Procedure Each subject was tested individually in a single session. The subject sat in a quiet room in front of an Apple IIe computer and monitor. After being instructed (in German), he or she was presented with a training block of stimuli from either the *ba-pa* series or the *ba-da* series. After each stimulus, the subject responded by pressing one of two keys. For the *ba-pa* series, the keys were labeled "*ba*" and "*pa*" whereas the second key was labeled "*da*" for the other series. Feedback was provided on the computer monitor during the training block.

Following the training block, the subjects were run on a test block for that series. No feedback was given during the test block. After completing the test block, the subjects were given a short rest before being tested on a training block and test block with the other series.

In summary, the first half of the experiment consisted of four blocks of trials, two training blocks and two test blocks. The second half of the experiment consisted of a second test block with each series. There was a lengthy pause between the two halves during which the subjects rested and performed in a different experiment. Half of the subjects were tested on the *ba-pa* series first and half were first tested on the *ba-da* series. The order was reversed for the second half of the experiment.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

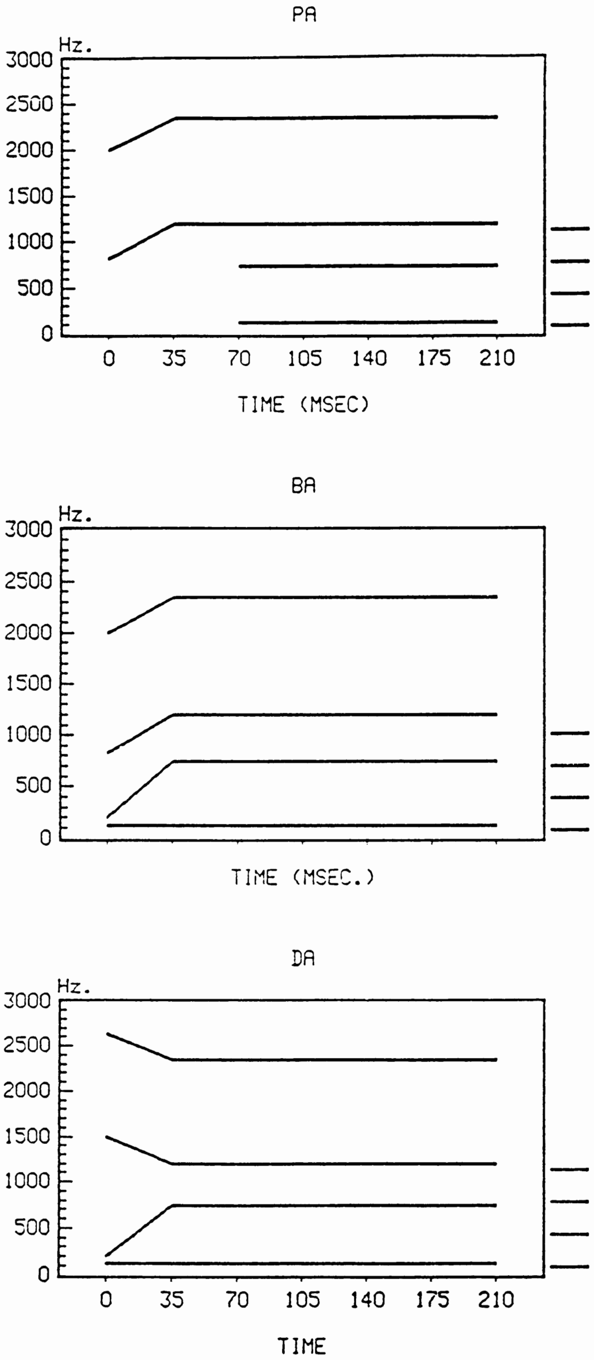


Figure 31.2 Primary formant trajectories for synthetic */ba/*, */da/*, and */pa/* syllables.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

Results and Discussion

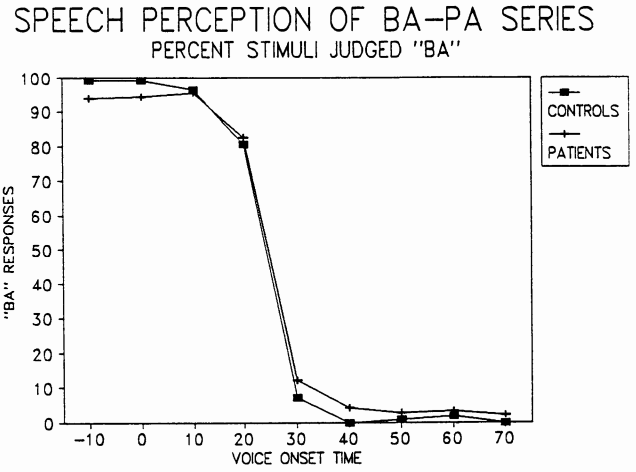
The mean number of *ba* responses was calculated for each of the nine stimuli from each continuum. One transformation was applied to the data before the averages were obtained. It took into account the fact that the point at which there is a shift from hearing *ba* to hearing either *pa* or *da* differs across subjects. For example, on the *ba-pa* series, six of the seven control subjects switched from *ba* responses to *pa* responses between stimulus 4 and stimulus 5. The 50 percent crossover point for the remaining subject was between stimulus 5 and stimulus 6. Similarly, the majority of responses to stimulus 4, VOT = 30 ms, were "*pa*" for one of the patients. Given these differences, the group functions would not appear as steep if the raw data were averaged across subjects without adjusting for individual differences in the crossover point. Thus, the data were realigned so that the crossover points for all of the subjects were the same stimulus. The crossover point could be unambiguously identified for all subjects. Note that by shifting the function in one direction, no entry was available for one of the tails. To correct this, a perfect score was entered there since subjects rarely made errors on the extreme stimuli.

This transformation was only required for the two subjects mentioned above on the *ba-pa* series. However, on the *ba-da* series, transformations were required for two control subjects and seven of the cerebellar patients. These centered the crossover point between stimulus 6 and stimulus 7. The direction of the transformation was to the right for the two control subjects and three of the patients. The remaining four patients required a leftward transformation. A transformation to the right indicates that the subject responded *da* more than average, whereas a transformation to the left indicates a relative bias to respond *ba*. The fact that so many subjects required transformations for this series points to a general problem encountered with the *ba-da* series. The stimuli in this series were more likely to be labeled *ba* than *da*. We chose our stimulus values on the basis of past research with American subjects (Reed 1984). These values may not have been ideal for German subjects, since the optimal stimulus set should lead to an equal distribution of responses between the two candidates.

Identification functions derived from the mean data for the cerebellar patients and control subjects are presented in figure 31.3. Results for the *ba-pa* series appear in the top half and results for the *ba-da* series appear on the bottom half. The functions indicate that the stimuli were generally perceived categorically. Qualitatively, the transition from one response category to the other appears similar to that obtained in previous studies of categorical perception in speech (e.g., Abramson and Lisker 1968). Within each category there is little change in performance.

For the *ba-pa* series, the patients tended to make a few more errors on each stimulus level. To assess this outcome further, we fit each subject's data with a normal ogive function and calculated the mean and standard deviation (Kling and Engen 1971). The mean of the standard deviations for each group on the two series appear in the top half of figure 31.4.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user



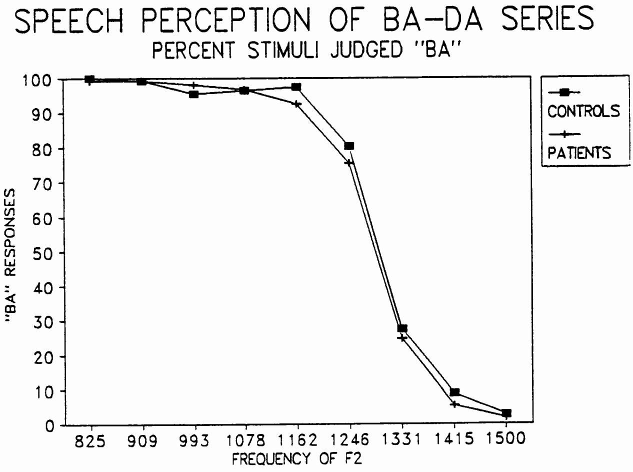


Figure 31.3 Identification functions on the *ba-pa* and *ba-da* continua for the cerebellar patients and control subjects.

These data were entered into a 2 × 2 ANOVA with one between-subjects factor (*group*: normal versus patient) and one within-subjects factor (*syllable*: *ba-pa* versus *ba-da*). A marginally significant effect of *syllable* emerged ($F(1, 20) = 4.08$, $p = .057$). Standard deviation estimates were higher for the *ba-da* continuum. However, the interaction depicted in figure 31.4 was not significant ($F(1, 20) = 2.29$, $p = .15$).

Although the preceding statistical procedure has been used to examine data in other experiments on categorical perception (e.g., Reed 1984), it may not be appropriate for theoretical and practical reasons. First, the data are fit to a normal ogive, the function assumed to describe most psychophysical data. However, functions interpreted as indicative of categorical perception are, by

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

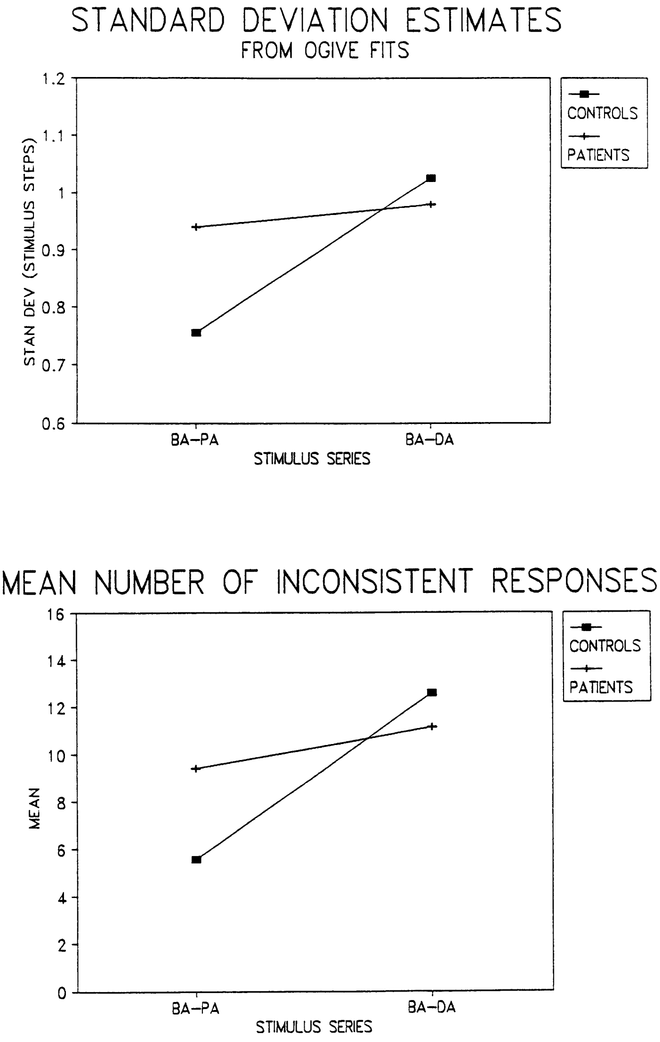


Figure 31.4 *Top:* Mean of the standard deviations for each subject group on the two continua. *Bottom:* Mean number of inconsistent responses made by each subject group on the two continua.

definition, poorly described by this function. Second, the estimation procedure requires that the raw data be transformed to *z*-scores. A decision must be made about what *z*-score to use when all sixteen presentations of a stimulus are assigned to the same category. We used a *z*-score of ±3.0 for such cases in the above analysis. Different values would alter the statistical outcome, although the interaction terms do not achieve significance when *z*-scores of either 2.2 or 4.0 are used for the perfect scores.

Given these problems, we designed a new measure of response inconsistency for each stimulus labeling function. Here, the total number of inconsistent responses was tallied. For example, if the number of *ba* responses by a subject for the nine stimuli of the *ba-pa* series was 16-16-16-15-13-1-0-0-0,

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

then the number of inconsistent responses would be $0 + 0 + 0 + 1 + 3 + 1 + 0 + 0 + 0 = 5$. Two inconsistency scores were calculated for each subject, one per series.

The means of the inconsistency scores appear in the bottom half of figure 31.4. An analysis of them yielded results similar to those in the previous ANOVA. More inconsistent responses were obtained with the *ba-da* series ($F(1, 20) = 4.23$, $p = 0.053$), perhaps reflecting the greater similarity between /*ba*/ and /*da*/ than between /*ba*/ and /*pa*/ (Shepard 1980; Miller and Nicely 1955). Most important, there was no difference between the patients and controls ($F(1, 20) < 1.0$) and the syllable by group interaction was not significant ($F(1, 20) < 1.54$, $p = .23$).[1]

In summary, the evidence does not indicate that the cerebellar patients have a deficit in identifying speech syllables that differ in VOT. While the means in figure 31.4 show that the patients tended to be more variable in identifying the VOT continuum and performed slightly better than the control subjects on the frequency continuum, the interaction was not significant (but see note 1). Further research will be needed to determine if a subset of cerebellar patients are impaired in categorizing stimuli that vary in VOT.[2] For example, the four subjects who performed most inconsistently on the VOT task also produced poor scores on a duration-perception task with pure-tone stimuli (Ivry and Keele 1989). However, some of the patients who were perfect on the categorical-perception task were impaired on the time-perception task. Thus, at present, we favor the conservative interpretation that the psychological mechanisms involved in the categorical perception of speech do not utilize the cerebellar timing system.

31.2 EXPERIMENT 2

From the preceding experiment, we tentatively conclude that the cerebellar timing system is not involved in one speech perception task that might require a temporal computation, the perception of syllables differing in VOT. We cannot differentiate between two interpretations of our null results. One interpretation is that the neural systems involved in language are independent of the cerebellar timing system, a variant of vertical modularity. Alternatively, the essential cues underlying VOT perception may not require timing (see Miller, Weir, Pastore, Kelly, and Dooling 1976). In experiment 2, we further assess the role of the cerebellum in language processes by investigating whether the cerebellar timing system plays a role in speech production. Specifically, can the cerebellar timing hypothesis account for speech disorders associated with dysfunction of this neural structure?

Patients with cerebellar disorders frequently suffer from speech dysarthria. Their symptoms include irregular articulation, fluctuating pitch, irregular rate and stress, and erratic phonation (Darley, Aronson, and Brown 1969; Kluin et al. 1988). Such disturbances have been characterized as an ataxic dysarthria rather than a spastic dysarthria (Darley, Aronson, and Brown 1969; Kluin et al.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

1988). Ataxia is generally associated with voluntary movements, whereas spasticity primarily reflects disturbances of posture.

Many deficits in the control of arm movements produced by patients with cerebellar lesions can be explained by the timing hypothesis. For instance, hypermetria and intentional tremor may reflect the inability to anticipate when the antagonist muscle needs to be activated to terminate a movement properly (see Ivry and Keele 1989). In experiment 2, we explore whether the timing hypothesis provides a useful description of cerebellar speech dysarthria. Specifically, we examine whether patients with speech dysarthria have difficulty in coordinating the articulatory actions that determine whether a consonant is voiced or voiceless.

Method

Subjects Six male patients with diffuse cerebellar lesions and four age-matched control subjects (three males and one female) were tested. The subjects were all native German speakers, between the ages of 49 and 60. All of the patients had cerebellar atrophy, and all were identified by clinical examination as presenting symptoms of speech dysarthria. Four of them and two of the control subjects had participated in the perception tasks of experiment 1. Two new patients and two new control subjects were recruited.

Procedure There were two different protocols used to obtain the speech samples. The subjects in protocol 1 were asked to alternate between saying the syllables *ba* and *pa*. The experimenter provided a pacing signal by moving his arm up and down. The syllables were produced at a rate of approximately one every 850 ms. Each trial contained between 12 and 15 alternations of the two syllables, yielding a total of 24 to 30 responses. If the subject became confused about which syllable was next in the sequence, the experimenter terminated the trial and began the procedure again. Subjects in protocol 1 were also tested on trials where they either repeated a single syllable or alternated between *ba* and *da*. The analysis of these trials is not included here.

Subjects in protocol 2 were tested more systematically. During testing, they sat in front of a computer monitor. Every four seconds, the computer generated a 50-ms warning tone, followed 500 ms later by the presentation of one of four printed nonsense syllables: *ba*, *pa*, *da*, or *ta*. The subjects were instructed to read the syllable in a normal voice. After the subject said the syllable, the experimenter triggered the computer to continue, and one second later the next trial began. On average, the subjects produced one syllable every 3.5 seconds.

Protocol 2 has two advantages over protocol 1. First, there was no rhythmicity to the subjects' responses and thus nonlinguistic temporal deficits should not be relevant. Second, since the subjects were presented with a target syllable, we could directly verify if their responses matched the stimuli.

The speech samples were recorded on high-quality metal cassette tapes. Subjects were tested in a quiet laboratory room.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

Data Analysis We analyzed all of the speech samples with the Interactive Laboratory Systems (ILS) software (Signal Technology, Inc.). The digitizing hardware, however, differed for the two protocols. For protocol 1, the output from the cassette recorder was low-pass filtered at 4.8 kHz and then sampled by a 12-bit analog-to-digital converter. For protocol 2, the recorder output was low-pass filtered at 9.6 kHz before conversion. The sampling rate for both protocols was set at 20 khz and the amplitude of the signal output was adjusted so that it spanned at least half the range of the A-D converter.

Analysis of Voice-Onset Time The VOT values of subjects' responses were determined from a visual display of the acoustic waveform. The judges, one for protocol 1 and one for protocol 2, were unfamiliar with the purpose of the experiment. On each trial, the judge located a syllable and marked two points. The first deviation from the background signal was identified as the onset of the burst of the stop consonant. Then, the judge identified the point where clear periodicity in the signal was evident. This was taken as the onset of voicing. The difference between these two points was recorded as the VOT for the syllable. Both authors verified the judges' performance for a subset of the trials.

For protocol 1, the data were displayed with a temporal resolution of 1 pixel per 6.4 ms, constraining the recorded VOTs for each trial to be multiples of 6.4 ms. Response syllables were selected in random order for analysis, rather than being evaluated in the order spoken. Thus, the judge did not know the hypothesis being tested, the subject's group (patient or control), or the syllable being evaluated.

For protocol 2, the temporal resolution of the displays was increased to 1 ms. Examples of three response-waveform displays are presented in figure 31.5. The judge heard and identified each spoken syllable before determining its VOT onset and offset. He was naive about the purpose of the experiment and did not know the subject's neurological status. Nor did he know the typical VOTs of the four syllables.

Spectral Analysis The judge for protocol 2 also performed a spectral analysis of the subjects' responses. In the waveform displays, he identified the midpoints of the vocalic portions of the syllables, defined as the points where their amplitudes were greatest. The formant frequencies around these points were then calculated by the ILS software. Frequency values were obtained for F1 and F2. These formants correlate roughly with the up-down and front-back positions of the articulators, respectively (Pickett 1980).

The spectral analysis was performed with a window spanning 5-ms intervals. Because data were sometimes missing from this analysis, an algorithm was used to calculate the frequencies for each trial. As part of it, F1 and F2 values were examined over a 30-ms epoch, spanning ±15 ms around the midpoint. If values of F1 and F2 were present for at least four of the seven measured points (−15 ms to +15 ms in steps of 5 ms), then the average of these values was recorded for that syllable. If there were values present for

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

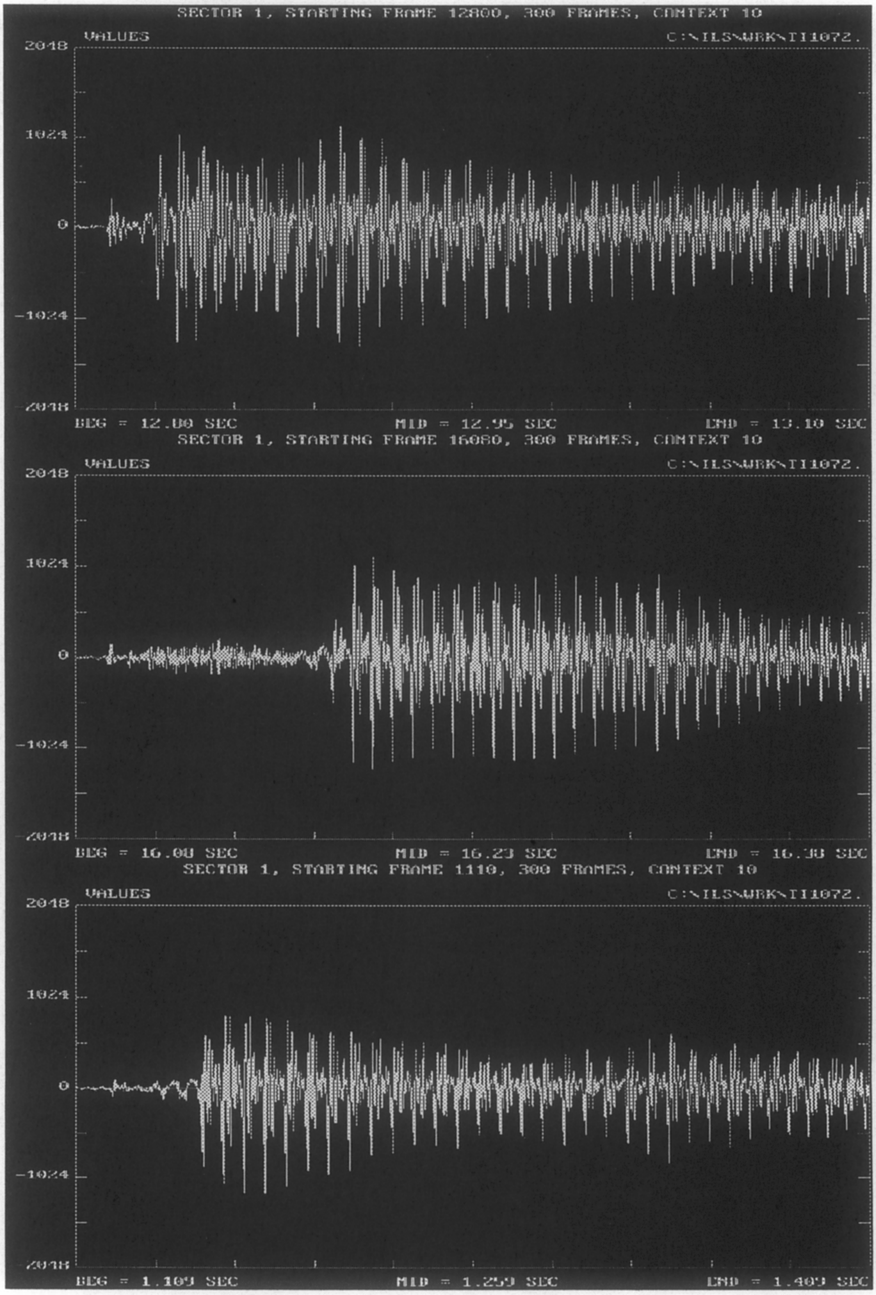


Figure 31.5 Examples of waveforms from which VOTs and vowel midpoints were calculated. Three syllables produced by patient B.U.C. The top trace is *da* with VOT = 9 msec. The middle trace is *pa* with VOT = 72 msec. The bottom trace is *pa* with VOT = 16 msec. All three traces are 300 msec in duration with burst onset at approximately 12 msec. Criterion for identifying onset of voicing was the first downward minimum of the periodic portion of the waveform. In all three examples shown, this point occurred one or two cycles prior to the large increase in signal amplitude.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

fewer than four of the measured points, frequency data were not recorded. Spectral analyses were only performed on male speakers.

The VOT analysis provides an acoustic measure of articulatory events involved in the coordination between a consonant and a vowel. We hypothesize that timing control is essential during this dynamic phase. In contrast, we hypothesize that precise timing is less important during the relatively static portions of speech such as the steady-state segment of vowels. At normal speaking rates, the articulators maintain a stable configuration, however briefly, during the production of vowels; the need for timing control may be reduced during this relatively static phase of speech.

To summarize, we hypothesized that the cerebellar patients would be highly variable during the dynamic phases of speech. This should be observed in the VOT measure which presumably reflects the coordination of muscles involved in voicing and release of the consonantal burst. On the other hand, we expect that the patients will be relatively less variable during the steady-state portion of vowel production because timing demands are reduced there. This reduced variability should be revealed by the spectral analysis.

Results and Discussion

Protocol 1 The numbers of obtained VOT estimates differed across subjects because of variation in the quality of the recordings. Approximately ten productions of each syllable were analyzed per subject. The resulting data were averaged separately over the two control subjects and four patients.

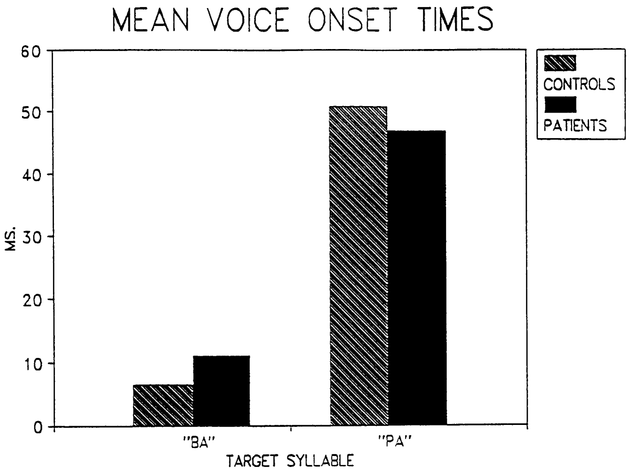
The top half of figure 31.6 presents the mean VOTs for the *ba* and *pa* syllables when produced alternately. There was a large difference in the mean VOTs for the productions of the voiced *ba* and voiceless *pa*. As in previous research (Lisker and Abramson 1964), no overlap of VOTs occurred between the two distributions for the control subjects.

The bottom half of figure 31.6 presents the means of the individual standard deviations. The patients were much more variable in their VOTs. Given that the mean VOTs of the patients and control subjects were quite similar, especially for the voiceless *pa*, it appears that the patients produced VOTs both shorter and longer than the control subjects. For example, the measured VOTs of *pa* syllables produced by one patient ranged from 6.4 ms to 205 ms. The VOT variability was perceptually salient with the lower end of the continuum sounding like *ba* and the upper end like a *pa* with a loud, breathy burst.

All four patients produced at least one *pa* that sounded like *ba*. Two of them also produced at least one *ba* that sounded like *pa*. However, it should be noted that these auditory judgments of the German subjects' responses were made by American listeners.

Given the small number of subjects and violations of homogeneity of variance, statistical analyses were not performed. Nonetheless, the finding of increased variability in the patients was consistent. For example, the standard deviation of the voiceless syllables produced by the most consistent patient

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user



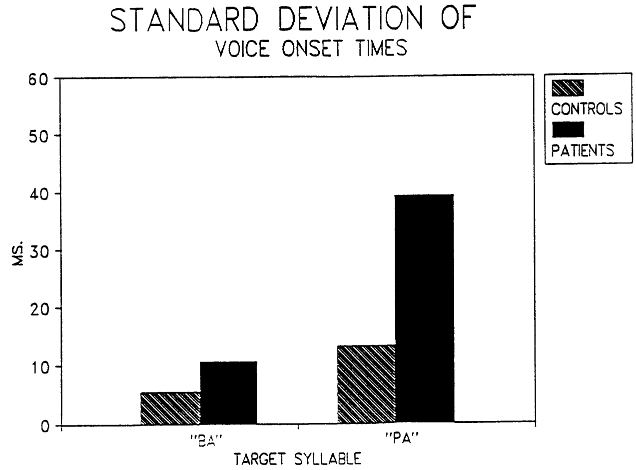


Figure 31.6 *Top:* Mean VOTs for the *ba* and *pa* syllables from protocol 1. *Bottom:* Mean standard deviations of the VOTs.

was 21.2 ms, whereas the comparable values for the control subjects were 12.0 and 14.3 ms.

Protocol 2, VOT Analysis The preceding analysis indicates that cerebellar patients are more variable in producing the articulatory actions that underlie VOT. However, as noted earlier, there are two problems with the previous data. First, the resolution of the VOT measurement was only 6.4 ms. Second, protocol 1 does not provide a clear separation of errors that resulted from selection of the wrong goal and errors that occurred in the process of achieving that goal. For example, in trying to say *pa* while alternating between *ba* and *pa,* a token may have a short VOT. This could reflect selection of the

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

wrong syllable to produce (*ba* instead of *pa*) or incorrect articulation of *pa*. Protocol 2 avoids these problems. Here, the measurement resolution was 1 ms and all of the stimuli were presented explicitly with an intertrial interval of approximately 3.5 seconds.

Figure 31.7 presents the mean VOTs and standard deviations for the patient and control groups. For simplification, we have combined the results over place of articulation: the results for *ba* and *da* were averaged together, as were the results for *pa* and *ta*. Thus, each data point represents approximately 48 productions per subject. With one exception, the mean values agree closely with the data in figure 31.6. The exception is the mean VOT for the control subjects on the voiceless syllables. The mean VOT was 43.5 ms for one control

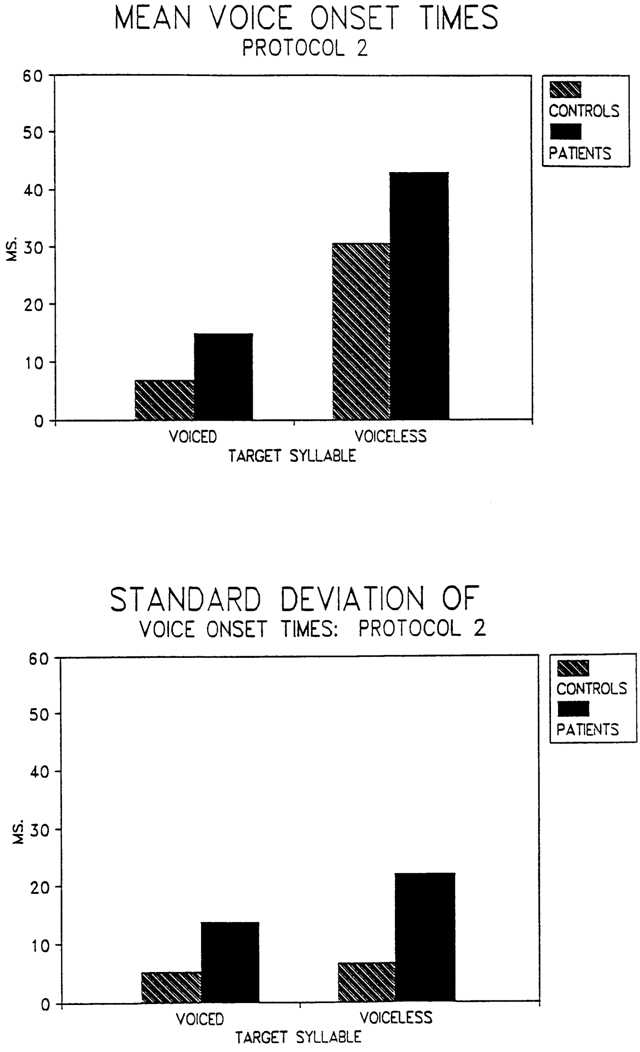


Figure 31.7 *Top:* Mean VOTs for the *ba* and *pa* syllables from protocol 2. *Bottom:* Mean standard deviations of the VOTs.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

subject, but was only 17.8 ms for the other subject. Despite the short VOT, the productions were almost always perceived correctly by the rater.

As in protocol 1, the patients were more variable than the control subjects on both the voiced and the voiceless syllables. To analyze these results statistically, we performed a series of pairwise tests of homogeneity of variance, assessing the null hypothesis that the variability observed for a patient equaled the variability observed for a control subject. Rather than use standard F ratios, we used the q-statistic discussed by Zar (1984, 204). This provides a conservative test for comparing variances in which the overall Type I error probability is held constant by adjusting the critical value of q as a function of how many pairwise comparisons are involved. A q-value is calculated as

$$q = (\ln(\mathrm{Var}_{patient}) - \ln(\mathrm{Var}_{control}))/sqrt(2/n_{patient} + 2/n_{control})$$

when experimental conditions are compared to control conditions.

The q-values derived from these pairwise comparisons demonstrate the consistency of the differences in VOT variances between the two patients and two control subjects (table 31.1). All of the patients' variances were significantly greater than those of the control subjects at the .05 level.

The variance differences in the acoustic signals were also apparent perceptually. As the subjects' responses were analyzed, the judge tried to classify them. The misclassifications are summarized in table 31.2. Very few errors occurred for the control subjects. In contrast, 13 percent of the patients' responses were misclassified (26 errors divided by 200 productions), almost all of which involved confusions along the temporal dimension. The errors were as likely to involve a voiced syllable being labeled voiceless as the reverse. This result was obtained for both of the cerebellar patients.

Two final points concern the error data. First, there were no errors for either group in which the perceived syllable differed from the stimulus on both the place and VOT features. Second, to repeat our earlier caveat, the judge was a native American-English speaker whereas the patients were German. However, any language-specific artifacts would be apparent in the classifications of responses by both the controls and patients.

Protocol 2, Spectral Analysis The vowel midpoint occurred approximately 100 ms after the release of the voiced consonants and was slightly later for

Table 31.1 q Values from Pairwise Comparisons of Voice-Onset-Time Variances in Protocol 2

	Voiced Consonants		Voiceless Consonants	
Control	WER	SCC	WER	SCC
Patient				
BEN	7.72*	9.21*	7.14*	6.30*
BUC	3.77*	5.11*	10.14*	9.29*

* $p < .01$ (Critical value = 2.93 for eight comparisons; at alpha = 0.5, critical value = 2.34.)

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

Table 31.2 Errors in Perceptual Classifications from Protocol 2

	Stimulus	Response	Total
Controls			
	voiced	voiceless	2
	voiceless	voiced	0
		VOT errors	2
	labial	alveolar	3
	alveolar	labial	1
		place errors	4
Patients			
	voiced	voiceless	12
	voiceless	voiced	11
		VOT errors	23
	labial	alveolar	0
	alveolar	labial	3
		place errors	3

the voiceless consonants (98 ms versus 120 ms for voiced and voiceless consonants, respectively). Although the vowel was the same in all of the syllables, consistent differences were found between the labial and alveolar syllables: F1 was slightly lower for the alveolars than the labials (mean = 690 hz versus 707 hz) and F2 was higher for the alveolars (mean = 1280 hz versus 1173 hz). Thus, in the following analysis, we combined the results from *ba* and *pa* into one group, and *da* and *ta* into another.

The mean formant frequencies fell within the normal range for male speakers. More informative are the mean standard deviations of F1 and F2, which appear in the top and bottom halves of figure 31.8, respectively. As the figure shows, the formants produced by the patients tended to be more variable across trials.

To test the reliability of this result, we again conducted a series of pairwise comparisons on the spectral data using the q-statistic mentioned previously. As shown in table 31.3, the difference between the patients and control subjects for F1 and F2 variability was inconsistent. Nine of the sixteen pairwise comparisons were significant; the remaining seven were not. Moreover, the median q value for the nine significant comparisons involving the formant frequencies was 4.31, considerably lower than the median q value of 7.43 obtained for the comparisons involving VOTs (table 31.1).

In summary, patients with cerebellar dysarthria produced syllables with greater variability in VOTs. This effect occurred for both voiced and voiceless syllables produced alternately and in isolation. We believe that these results reflect a difficulty in temporally coordinating the neuromuscular events responsible for voicing and release of the stop consonant. This hypothesis is consistent with the cerebellar timing hypothesis.

However, other interpretations of these data are possible and cannot be ruled out given our current analyses. For example, VOT varies as a function

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

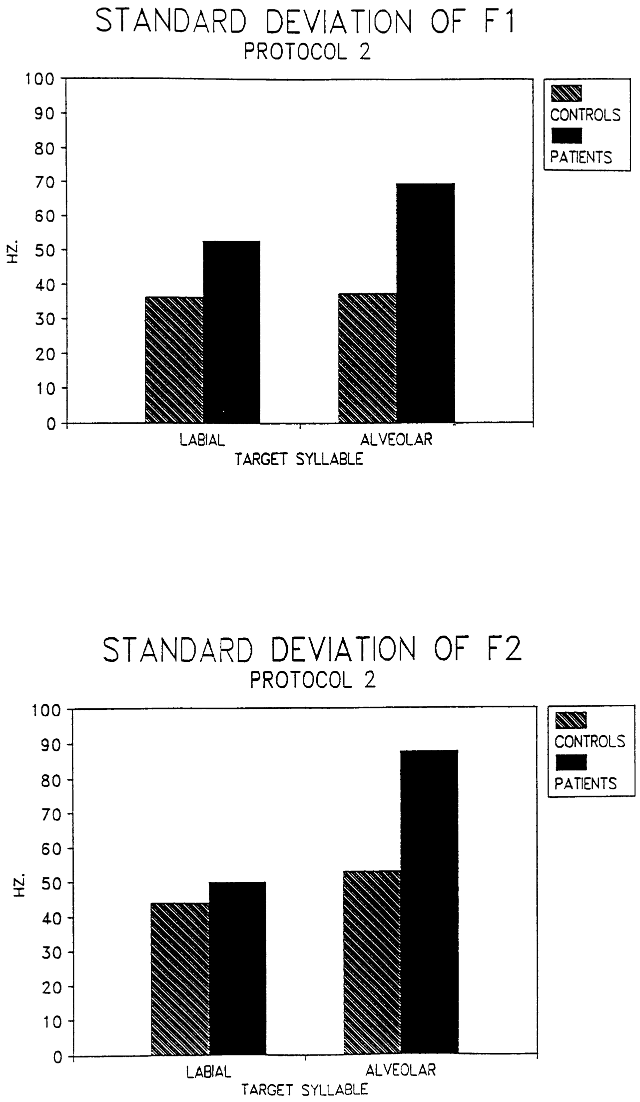


Figure 31.8 *Top:* Mean standard deviation of F1 at vowel midpoint from protocol 2. *Bottom:* Same for F2.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

Table 31.3 q Values from Pairwise Comparisons of First and Second Formant Steady-State Frequency Measures in Protocol 2

F1	Labial Consonants		Alveolar Consonants	
Control	WER	SCC	WER	SCC
Patient				
BEN	5.41*	4.65*	6.74*	3.93*
BUC	−2.21	−2.37	4.30*	1.29

F2	Labial Consonants		Alveolar Consonants	
Control	WER	SCC	WER	SCC
Patient				
BEN	3.47*	−1.25	4.31*	1.95
BUC	3.99*	−1.00	4.62*	2.24

* $p < .01$ (Critical value = 2.93 for eight comparisons/meausre; at alpha = 0.5, critical value = 2.34.)

of speaking rate (Miller, Green, and Reeves 1986). Thus, the increased VOT variability may arise indirectly as a consequence of greater variability in terms of overall speaking rate. If the syllable durations were more variable for the patients' productions, it would then be necessary to compute a relative VOT measure that allows an adjustment for the duration of the individual syllable tokens. Syllable duration measurements were not made on the current data set, but will be obtained in future studies. This should allow stronger tests of whether the increased variability is manifest at the level of VOT, syllable duration, or both. In other words, at what level are speech deficits in patients with cerebellar lesions manifest—phonemic, syllabic, or suprasegmental? Moreover, if articulatory deficits are identified for linguistic phenomena spanning different temporal ranges, it will be of interest to determine whether the timing hypothesis can account for all or just a subset of these deficits.

The spectral analysis performed here (fig. 31.8 and table 31.3) was included as an initial attempt to obtain an acoustic measure independent of timing control. We anticipate that vowel formant frequencies may reflect the configuration of the articulators during steady-state speech. We hypothesized that a loss in timing control would be more marked during the dynamic period characteristic of consonant articulation. The results of experiment 2 provide mixed support for this prediction. The patients were generally more variable than the control subjects on the spectral measures, but this deficit was less consistent and quantitatively less robust than that observed in the production of VOT. The distinction between static and dynamic articulatory events may prove useful to researchers examining the underlying bases of different speech disorders.

31.3 EXPERIMENT 3

In what follows, we briefly present two additional experiments that explore a different temporal regularity in language. In certain written languages, word

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

Table 31.4 Sample Stimuli for Gemination Perception Experiment

	Stimulus Condition	
	Geminates	Controls
Examples:	seno/senno (breast/wisdom)	seno/seco (breast/with)
	seco/secco (with/arid)	secco/senno (arid, wisdom)
	sete/sette (thirst/seven)	sete/sere (thirst/evenings)

pairs exist in which the only orthographic difference is that the medial consonant is doubled for one member. This phenomenon, known as gemination, is found in Italian. Examples of Italian geminate pairs appear in table 31.4. As can be seen from them, there is no obvious connection between the word meanings within each pair. While the distinction between the two members of a pair is obvious orthographically, a different (acoustic) cue must be used to discriminate them in oral language.

One way that they are differentiated is to lengthen the duration of the word with the doubled consonant (referred to as the doublet). This lengthening occurs in the medial consonant (and sometimes in the vowel preceding this consonant). With some pairs, the difference in duration is the only salient perceptual cue; other pairs may differ on additional dimensions such as syllabic stress or vowel pronunciation.

As with categorical perception, the temporal difference in the acoustic signal between members of a geminate pair may or may not be directly exploited by psychological mechanisms. In experiment 3, we investigate whether patients with cerebellar lesions have difficulty discriminating between geminate-pair members that differ in duration.

Method

Subjects Six patients, four men and two women (mean age = 44.2 years, SD = 10.5), with atrophy of the cerebellum were tested. The patients included a family pair of mother and son. One of the patients, the mother of the familial pair, showed symptoms indicative of olivopontocerebellar atrophy. All of the patients were native Italian speakers. Four of the patients were born and lived within 100 km of Rome. Another was from Naples and one was from northern Italy, a region where gemination is less marked.

Two men and one woman served as control subjects (mean age = 38.3 years, SD = 14.4). Two of them were native Romans, and one was a native Venetian. In addition, two non-Italian control subjects were tested, a 40-year-old man born in Chile, and a 32-year-old Brazilian male. Both of these subjects were fluent in conversational Italian, although it was not their native language.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

The subjects were tested at the Neurology Department of Catholic University in Rome.

Five of the patients and all of the Italian control subjects were tested on a version of the time-perception task reported in Ivry and Keele (1989). The mean standard deviations for the patients and control subjects were 74.4 ms and 29.3 ms, respectively, for a base interval of 600 ms. This difference is significant ($t(6) = 2.36$, $p < .05$), replicating our previous findings that cerebellar patients are impaired on the time-perception task.

Stimuli An Italian assistant, native to Rome, generated the pairs of geminate words for the experiment. During the course of two one-hour sessions, she produced 53 pairs (106 words). Twenty-eight of the stimulus words were digitized to verify that they embodied a difference in duration between the members of the geminate pairs. Non-negligible differences in duration occurred for the medial consonant and the vowel preceding the medial consonant (defined as the interval from vowel offset to vowel onset). Across the 14 pairs, the consonant was longer by 126 ms (SD = 58 ms) on average for the doublet. While the preceding vowel tended to differ between the two members of each pair, the direction of this difference was inconsistent: in half the cases, the vowel for the doublet was longer; in the other half, it was shorter.

In addition to the geminates, control words were selected for 76 of the geminates. These words differed from the corresponding geminates in that their medial consonants were changed. Here the differences were not temporal, but involved consonants that differed either in place, manner, or voicing of articulation or a combination of these three features. Some of the control words were also geminates, and some were new words (see table 31.4).

Through an oversight, one geminate word was omitted from the test stimuli, and thus the final stimulus set consisted of 181 words. These words were recorded individually on a cassette tape in random order, with an interstimulus interval of approximately 4 seconds. The set was divided into five blocks of between 35 and 38 words each. The words were spoken by the Italian assistant.

Procedure The subjects sat in front of a computer monitor. On each trial, a pair of words was displayed on the monitor. Then, a spoken stimulus, representing one of these words, was played from the tape recorder. The subject pressed one of two keys to indicate which word on the screen corresponded to the spoken stimulus.

There were 105 trials on which the displayed version of the spoken word and the alternative word formed a geminate pair. The two displayed words had different medial consonants on the remaining 76 trials. The spoken stimulus came from a geminate pair on half of the control trials and from the set of control words on the other half.

A short break occurred between test blocks.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

Results and Discussion

As noted earlier, in addition to duration, spoken members of geminate pairs may differ from each other in terms of other acoustic cues such as syllabic stress or vowel pronunciation. However, a decomposition of the geminate pairs into two groups, those with non-temporal cues and those in which the only salient cue was medial consonant duration, showed no difference in performance. Thus, the trials involving geminate pairs were merged into a single condition.

The percentage of erroneous classifications is presented in table 31.5. Cerebellar patients did not have a selective deficit in discriminating between members of a geminate pair. While the error rate was higher on these trials for the patients than for the control subjects, the same pattern also occurred on the control pairs. Indeed, there were very few errors in this task. If the data from the cerebellar patient with olivopontocerebellar atrophy are excluded, the percentage of errors on the geminate pairs drops to 1.4 percent.

The two nonnative control speakers made errors on 16.5 percent of the geminate pairs. In contrast, one of these subjects was perfect on the control pairs, while the other made 7 percent errors, yielding a mean of 4.7 percent. This suggests that the gemination discrimination is perceptually difficult, at least for nonnative speakers. Nonetheless, the patients were generally successful at this discrimination.

Experiment 3 provides a second source of evidence indicating that the integrity of the cerebellum is not critical for the perception of linguistic differences that can be described temporally. It must be noted that, for both experiments 1 and 3, this conclusion is based on null results, the lack of any performance difference between the patients and control subjects. However, the same cerebellar patients producing these null results were impaired on the time-perception task with nonspeech, tone stimuli. These results suggest that the requisite psychological processes for perceiving differences in voice-onset time or discriminating between members of a geminate pair do not include the cerebellar timing system.

31.4 EXPERIMENT 4

In our final experiment, we examine the production side of the gemination phenomena. The issue here concerns whether cerebellar patients can produce the acoustic cues that underlie perceptual discriminations between the mem-

Table 31.5 Percent Errors in Geminate Perception Experiment

Subject Group	Geminate Pairs	Control Pairs
Controls	1.3	2.1
Cerebellars	4.4	6.2
Nonnative speakers	16.5	4.5

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

bers of geminate word pairs. The results of experiment 2 indicated that patients with cerebellar lesions were impaired in coordinating the articulatory gestures that produce consistent differences between voiced and voiceless consonants. This production deficit can be interpreted as a deficit in temporally coupling the muscular actions needed for consonantal release and voicing. In experiment 4, we look for converging evidence that cerebellar dysarthria can be characterized as a problem in temporally controlling articulatory actions.

Method

Subjects The six patients with cerebellar atrophy and three native Italian control subjects from experiment 3 were tested. Given the difficulty in identifying patients with cerebellar lesions, we were unable to select only Italian patients with clinical evidence of speech dysarthria—the subjects who would be of most interest in a speech production study. When tested informally at the time of the experiment, four of the patients showed some signs of dysarthria. The speech of the other two patients appeared normal.

Stimuli The same 181 words were used.

Procedure Each subject again sat in front of the computer monitor. On each trial, a single word was presented on the monitor screen, and the subject read the word in a normal voice. The subjects' productions were recorded on cassette tape. The words were read in the same order as the stimulus words judged in experiment 3. There were five blocks of 35 to 38 words each.

Two new native Italian speakers were recruited to judge the taped productions. Each judge listened to each word that had been read, and then chose which of two alternatives it matched. Thus, the task for the judges was identical to the perception task of experiment 3, except that the judged words were produced by the patients and control subjects rather than by the Italian assistant.

As in experiment 3, on 105 trials the judged alternatives were geminate pairs, and on the other 76 trials the alternatives differed in terms of their medial consonants. The judges listened to the productions of all nine subjects. The order of listening to the subjects was random, and the judges did not know whether a subject was a patient or a control.

Results and Discussion

Except in one case, there was near-perfect agreement between the classifications of the two judges. The exception occurred for eight words in which the syllable stress was inconsistent across subjects. These words have been excluded from the analysis.

The top half of figure 31.9 depicts the mean percentage of errors in each condition for the two subject groups (patients versus controls). The figure indicates that, overall, the patients' syllables on both types of trials were more

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

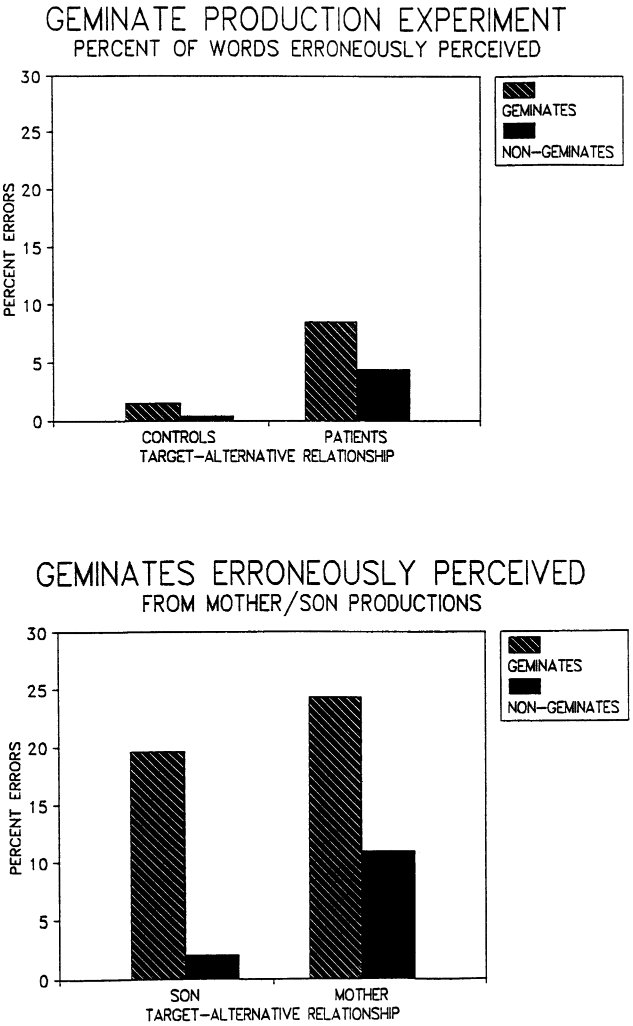


Figure 31.9 *Top:* Mean number of productions erroneously classified for each subject group on the geminate and control words. *Bottom:* Mean number of productions erroneously classified for two patients, a mother and son.

likely to be misperceived than the syllables produced by the control subjects. The interaction suggested in figure 31.9 may result from a floor effect because very few of the syllables produced by the control subjects were misperceived.

While errors were obtained for the syllables produced by all of the patients, over 60 percent of the perceptual discrimination errors were made in judging the productions of the mother-son patient pair. These two patients showed the most marked speech dysarthria of all the patients. The percentage of errors for these two subjects is presented in the bottom half of figure 31.9. Errors by the mother were obtained with both types of word pairs (geminates and non-geminates). In contrast, the son's deficit appeared to be restricted to the geminate pairs. Almost 20 percent of these pairs produced a misperception for

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

him whereas only 2 percent of the consonant pairs yielded errors. The mother's dysarthria and general disorder were much more marked.

These data raise an intriguing possibility: the articulatory actions that differentiate members of geminate-word pairs may be impaired during the early stages of cerebellar atrophy. This deficit perhaps reflects a reduced ability to temporally coordinate the requisite articulations. As the atrophy becomes more advanced, other articulatory deficits may emerge, perhaps because the atrophic process has spread to other neural systems. Further research is required to determine if the errors on the non-geminate trials can be explained by the timing hypothesis.

31.5 SUMMARY AND CONCLUSIONS

As reported in our previous research (Ivry et al. 1988; Ivry and Keele 1989; Ivry and Diener 1991), patients with cerebellar lesions are impaired on a variety of motor and perceptual tasks that require precise timing. These results have led us to propose that the cerebellum operates as an internal timing system. This hypothesis embraces the notion of horizontal modularity in that the same system is shared across task domains when temporal computations are needed. Such sharing constitutes a violation of vertical modularity.

Given that the cerebellar timing system is not restricted by vertical modularity, we are interested in determining the domain of this system. When are the temporal computations of the cerebellum needed? Over what temporal range do they apply? The experiments presented in this chapter provide our initial answers to whether timing phenomena in oral language involve the computational mechanisms of the cerebellum.

Two distinct linguistic phenomena have been studied here: the perception and production of voice-onset time in German, and the perception and production of gemination in Italian. There are obvious differences between these two linguistic phenomena including segment position (initial versus medial), and temporal range (VOT differences of approximately 40 ms versus geminate differences of 120 ms). Nonetheless, the results were similar. Cerebellar lesions did not produce consistent deficits in the perception of either VOT or gemination differences. Overall, VOT identification functions were similar for the patients and healthy subjects. Nor did the patients have a selective deficit in discriminating geminate pairs.

These null results suggest at least four plausible interpretations. First, although temporal features in the acoustic signal provide a salient means for describing voicing and gemination, the actual psychological processes may not exploit these cues, at least not in a direct manner. If the psychological mechanism does not require precise temporal computations, we would not expect the integrity of the cerebellum to be critical. Second, the psychological process may use the temporal information, but the process may not explicitly compute this information. Perhaps the temporal cue is derived implicitly. The cerebellar timing system may only be invoked for explicit temporal computations. Third, temporal information may be essential for VOT and gemination perception,

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

but the time differences involved in these tasks may not be in the range encompassed by the cerebellar timing system (see Keele and Ivry, in press). Other neural systems may be essential for short-duration phenomena, and these may be essential for processing the temporal information associated with VOT and gemination. Fourth, there may be a timing module independent of the cerebellar timing system that is specialized for speech perception (Liberman Mattingly 1985). Of course, these interpretations are not mutually exclusive. Some combination of them may be the actual case.

In contrast to the null results obtained from our perception experiments, the timing hypothesis provides an interpretation of cerebellar speech dysarthria in production. The VOTs produced by the patients of experiment 2 were more variable, and led to perceptual errors along the voicing dimension. Similarly, the Italian patients, especially the familial mother/son pair, frequently failed to produce an unambiguous distinction between the two members of a geminate pair. These production deficits may reflect an inability to temporally coordinate events across different sets of articulators (e.g., those involved in the onset of voicing and the release of consonantal airflow). This type of control would, of course, be most needed during the dynamic portions of speech when the articulators are rapidly changing their configuration.

If this hypothesis is correct, then the cerebellar contribution to speech production would be less critical during the relatively stable periods of articulation, namely during steady-state vocalic segments. Instead, the primary control problems during these periods are more configural or static, rather than dynamic. The articulators must achieve and maintain the proper configuration to create certain target resonances. Muscular events are maintained rather than dynamically coordinated.

Our first test of this corollary hypothesis in experiment 2 has been inconclusive. While the differences between the patients and controls seemed less robust, the vocalic formant frequencies produced by the patients were generally more variable than those produced by the control subjects. Perhaps there is not really a true steady state in speech; all articulatory events may be dynamic. Our future research will continue to explore the usefulness of a dynamic/static distinction, both as it applies to understanding the role of the cerebellum in motor control, and as a way to contrast different disorders of speech.

NOTES

This research was supported by an Office of Naval Research Contract (N00014-87-K-0279) and a fellowship from the Alfred P. Sloan Foundation. The authors thank Steven Keele for his help throughout this project as well as David Meyer, Sylvan Kornblum, Peter Gordon and Herbert Heuer, and Al Liberman for helpful comments. H. C. Diener and G. Macchi graciously provided facilities and patient access for the experiments. Lori Kennedy and Warren Branson provided invaluable help in performing the acoustic analyses, and Paul Lebby assisted in the statistical analyses.

1. The inconsistency measure weights all stimuli equally. However, it might be argued that errors more distant from the crossover point should receive greater weight. Thus, linear integer

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

weights were employed in an analysis, with errors on the crossover stimulus given a weight of 0.0, errors on stimuli neighboring the crossover point given a weight of 1.0, errors on stimuli 2 steps from the crossover a weight of 2.0, and so forth. An ANOVA with this measure revealed a significant interaction of group X *syllable* ($F(1, 20) = 5.56, p = 0.029$). The patients were more inconsistent when identifying the stimuli from the *ba-pa* continuum.

2. This hypothesis would predict that there should be more variability across patients in comparison to variability across the control subjects. A test of heterogeneity of variance was performed on the standard deviations of the group means from the analysis of the ogive fits. The group standard deviations were not significantly different from each other for the *ba-pa* series ($F(14, 6) = 2.16, p > .10$). A similar test for the *ba-da* series was not conducted since the standard deviation of the control group mean was larger than that obtained for the patients.

REFERENCES

Abbs, J., and Gracco, V. (1983). Sensorimotor actions in the control of multi-movement speech gestures. *Trends in Neuroscience, 6,* 391–395.

Abramson, A., and Lisker, L. (1968). Voice timing: Cross-language experiments in identification and discrimination. Paper presented at the meeting of the Acoustical Society of America. Ottawa.

Atkeson, C. (1989). Learning arm kinematics and dynamics. *Annual Review of Neuroscience, 12,* 157–183.

Darley, F., Aronson, A., and Brown, J. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12,* 246–269.

Flash, T. and Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience, 5,* 1688–1703.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Fowler, C. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8,* 113–133.

Gordon, P., and Meyer, D. (1984). Perceptual-motor processing of phonetic features in speech. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 153–178.

Hallett, M. and Khoshbin, 5. (1980). A physiological mechanism of bradykinesia. *Brain, 115,* 465–480.

Hasan, Z. (1986). Optimized movement trajectories and joint stiffness in unperturbed, inertially loaded movements. *Biological Cybernetics, 53,* 373–382.

House, A. (1961). On vowel duration in English. *Journal of the Acoustical Society of America, 33,* 1174–1178.

Ivry, R. (1990). Multi-task component analysis of movement and perception in normal and clinical populations. Paper presented at the annual meeting of North American Society for the Psychology of Sport and Physical Activity. Houston.

Ivry, R., and Diener, H. (1991). Impaired velocity perception in patients with lesions of the cerebellum. *Journal of Cognitive Neuroscience, 3,* 355–366.

Ivry, R., and Keele, S. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience, 1,* 136–152.

Ivry, R., Keele, S., and Diener, H. (1988). Dissociation of the lateral and medial cerebellum in movement timing and movement execution. *Experimental Brain Research, 73,* 167–180.

Keele, S., and Ivry, R. (1988). Modular analysis of timing in motor skill. In G. Bower (Ed.), *The Psychology of learning and motivation.* Volume 21, 183–228. New York: Academic Press.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

Keele, S., and Ivry, R. (1991). Does the cerebellum provide a common computation for diverse tasks: A timing hypothesis. In A. Diamond (Ed.), *Developmental and Neural Basis of Higher Cognitive Function*, 179–211. New York: Annals of the New York Academy of Sciences.

Keele, S., Ivry, R., and Pokorny, R. (1987). Force control and its relation to timing. *Journal of Motor Behavior, 19*, 96–114.

Keele, S., Pokorny, R., Corcos, D., and Ivry, R. (1985) . Do perception and motor production share common timing mechanisms? *Acta Psychologica, 60*, 173–193.

Kelso, J., Bateson, E., Saltzman, E., and Kay, B. (1985) . A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modelling. *Journal of the Acoustical Society of America, 77*, 266–280.

Kent, R., and Moll, K. (1975). Articulatory timing in selected consonant sequences. *Brain and Language, 2*, 304–323.

Kewley-Port, D. (1978). *Executive program to implement the Klatt software speech synthesizer.* (Progress Report 4, Research on Speech Perception.) Bloomington: Indiana University.

Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society, 67*, 971–995.

Kling, J., and Engen, T. (1971). *Experimental Psychology*. 3d ed. New York: Holt, Rinehart, and Winston.

Kluin, K., Gilman, S., Markel, D., Koeppe, R., Rosenthal, G., and Junck, L. (1988). Speech disorders in olivopontocerebellar atrophy correlate with positron emission tomography findings. *Annals of Neurology, 23*, 547–554.

Kozhenvnikov, V., and Chistovich, L. (1965). *Speech articulation and perception*, J.P.R.S. Washington D.C. #30543.

Lacquaniti, F., and Soechting, J. (1982). Coordination of arm and wrist motion during a reaching task. *Journal of Neuroscience, 2*, 399–408.

Liberman, A., and Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1–36.

Lisker, L., and Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384–422.

Lisker, L., and Abramson, A. (1967). Some effects of context on voice-onset time in English stops. *Language and Speech, 10*, 1–28.

Lundy-Ekman, L., Ivry, R., Keele, S., and Woollacott, M. (1991). Timing and force control deficits in clumsy children. *Journal of Cognitive Neuroscience, 3*, 370–377.

Miller, G., and Nicely, P. (1955). Analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27*, 338–353.

Miller, J., and Liberman, A. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception and Psychophysics, 25*, 457–465.

Miller, J., Green, K., and Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica, 43*, 106–115.

Miller, J., Weir, L., Pastore, R., Kelly, W., and Dooling, R. (1976). Discrimination and labelling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America, 60*, 410–417.

Pickett, J. (1980). *The sounds of speech communication*. Baltimore: University Park.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user

Port, R. (1981). Linguistic timing factors in combination. *Journal of the Acoustical Society of America, 69,* 263–274.

Reed, M. (1984). Temporal processing and speech perception in readinq disabilities. Unpublished doctoral dissertation, University of Oregon.

Shepard, R. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210,* 390–398.

Stelmach, G., Teasdale, N., Phillips, J., and Worringham, C. (1989). Force production characteristics in Parkinson's disease. *Experimental Brain Research, 76,* 165–172.

Strange, W., and Jenkins, J. (1978). Role of linguistic experience in the perception of speech. In R. Walk and H. Pick (Eds.), *Perception and experience.* New York: Plenum.

Studdert-Kennedy, M. (1976). Speech perception. In N. Lass (Ed.), *Contemporary issues in experimental phonetics,* 243–293. New York: Academic Press.

Tuller, B., and Kelso, J. (1990). Phase transitions in speech production and their perceptual consequences. In M. Jeannerod (Ed.), *Attention and performance XIII.* Hillsdale, NJ: Erlbaum.

Tuller, B., Kelso, J., and Harris, K. (1983). Converging evidence for the role of relative timing in speech. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 829–833.

Weismer, G. and Fennell, A. (1985). Constancy of (acoustic) relative timing measures in phrase-level utterances. *Journal of Acoustical Society of America, 78,* 49–57.

Zar, J. (1984). *Biostatistical analysis.* Englewood Cliffs, NJ: Prentice-Hall.

Downloaded from http://direct.mit.edu/books/chapter-pdf/279663/9780262290906_cbt.pdf
by UCB - University of California Berkeley user