Behavioral and Cognitive Neuroscience Reviews
http://bcn.sagepub.com/

Temporal Organization of "Internal Speech" As a Basis for Cerebellar Modulation of Cognitive Functions

Hermann Ackermann, Klaus Mathiak and Richard B. Ivry

Behav Cogn Neurosci Rev 2004 3: 14
DOI: 10.1177/1534582304263251

Temporal Organization of "Internal Speech" as a Basis for Cerebellar Modulation of Cognitive Functions

Hermann Ackermann
Klaus Mathiak
University of Tuebingen

Richard B. Ivry
University of California

The sequencing of smooth and rhythmically "sculptured" words and phrases at a speaker's habitual speech rate (4 Hz to 6 Hz) critically depends on the cerebellum. Besides overt performance, the cerebellum also seems to organize the syllabic structure of "auditory verbal imagery" or "internal speech"—that is, a prearticulatory but otherwise fully elaborated and temporally organized representation of verbal utterances. As a consequence, cerebellar disorders may compromise cognitive operations that involve a speech code, such as verbal working memory, or disrupt cognitive processes that encompass linguistic mediation. Besides the temporal organization of syllable strings at a prearticulatory level, cerebellar patients are impaired in speech perception tasks requiring the encoding of durational parameters of the acoustic signal. The hemodynamic responses associated with these two aspects of verbal-acoustic communication—internal speech and speech perception—were found to be organized along the rostro-caudal direction within paravermal aspects of the superior right cerebellar hemisphere. Those areas of the right cerebellar hemisphere thus might provide a common platform for the computation of temporal aspects of verbal utterances in the domains of both speech production and perception.

Key Words: cerebellum, cognition, speech motor control, speech perception

A classic tenet of clinical neurology, dating back to the early 19th century, considers the cerebellum exclusively devoted to motor functions, such as the control of posture and gait or the coordination of voluntary limb movements and vocal tract muscles during speech production (Dow & Moruzzi, 1958; Lechtenberg, 1993). More recent investigations have led to proposals of cerebellar involvement in a wide range of cognitive tasks, including spatial processing, higher order language functions, and executive operations associated with planning, set-shifting, or verbal fluency (for a review, see Ackermann & Daum, 2003). Furthermore, a large number of functional-imaging studies have found cerebellar activation to be associated with a diverse range of cognitive capabilities, such as attentional control, lexical access, episodic memory retrieval, and executive functions (for a review, see Cabeza & Nyberg, 2000).

Although a consensus seems to be emerging that the functional domain of the cerebellum is not limited to motor control and sensorimotor learning, a computational account of how this subcortical structure contributes to nonmotor functions remains to be determined (Dolan, 1998). Schmahmann and Sherman (1998) proposed that the cerebellum modulates the efficiency of processing in other neural regions rather than contributing directly to these cognitive procedures (see also Akshoomoff, Courchesne, & Townsend, 1997). In this view, the cerebellum might support generalized (supramodal) operations acting across a variety of motor and cognitive domains. Hallett and Grafman (1997) assumed a frontocerebellar network subserving the storage and organization of timed (temporally labeled) sequences of both movement patterns and mental oper-

---

Authors' Note: This work was supported by grants of the German Research Foundation (DFG-SFB 550, DFG-SPP "Sprachproduktion"). We thank Dr. Ingo Hertrich and Maike Borutta for helpful assistance and comments.

Behavioral and Cognitive Neuroscience Reviews
Volume 3 Number 1, March 2004 14-22
DOI: 10.1177/1534582304263251

ations. In a similar vein, Molinari, Leggio, and Silveri (1997) suggested that the cerebellum operates as a "controller," mediating sequential organization of the various subcomponents of complex cognitive tasks.

Production of verbal utterances requires, at various processing stages, the sequencing of linguistic data sets. Sound structures must be aligned during overt and "internal speech," a prearticulatory, but otherwise fully elaborated (parsed) and temporally organized, representation of verbal utterances (Levelt, 1989). Ivry and Fiez (2000; see also Ivry, Justus, & Middleton, 2001) suggested that some of the disrupted cognitive functions observed in patients with cerebellar disorders may result from an impairment in covert (inner) verbal rehearsal processes. Functional-imaging studies have also found activation of the cerebellum to be associated with working memory tasks (e.g., Fiez et al., 1996; Frackowiak, Friston, Frith, Dolan, & Mazziotta, 1997). These hemodynamic responses have been attributed to a covert (silent) articulatory rehearsal mechanism (articulatory loop; Baddeley, 1992), a subcomponent of the working memory system that supports the short-term maintenance of linguistic (speech-encoded) information (but see Chein & Fiez, 2001).

Our work on the role of the cerebellum in speech production and perception provides a basis to combine these suggestions into a unified concept; because the cerebellum subserves the temporal organization of syllable strings even at the level of internal speech, cerebellar disorders would be expected to compromise executive and memory functions that engage rehearsal mechanisms based on a speech code (Ivry & Fiez, 2000) or to disrupt linguistic "scaffolding" of various cognitive operations (Jackendoff, 1997).

MOTOR ASPECTS OF SPEECH PRODUCTION IN CEREBELLAR DISORDERS: ACOUSTIC AND KINEMATIC ANALYSES

Syllabic Timing

Cerebellar disorders may compromise the control of vocal tract muscles during speech production. Samples of spontaneous verbal utterances as well as text passage readings indicate a prominent slowing of speech tempo in ataxic dysarthria (for a review, see Ackermann & Hertrich, 2000). In addition, these patients exhibit a tendency toward isochronous syllable durations within utterances (e.g., Kent & Rosenbek, 1982). Modulation of syllable lengths represents a salient information-bearing aspect of speech output that contributes both to the meaning of a sentence (linguistic prosody) and to a speaker's emotional expression (affective prosody). First, duration is the major acoustic correlate of word accent in German; stress-bearing syllables are significantly longer than their unstressed cognates (see, e.g., Riecker, Wildgruber, Dogil, Grodd, & Ackermann, 2002). Second, syllable prolongation, in addition to pitch and loudness variations, can enhance the affective tone of an utterance. Lengthening of the initial vocalic segment in the utterance "How beaútiful" indicates a speaker's surprise. Third, syllable prolongation may help direct a listener's attention. For example, the marked vowel /e/ in "She delivered the message to Jénny, not Charles" emphasizes the recipient of the respective action.

Based on acoustic analyses from a case study, Kent and Rosenbek (1982) assumed that isochronous syllable durations within utterances subsequent to cerebellar pathology reflect the prolongation of short vocalic elements compared to their long cognates. To obtain a comprehensive and systematic data set of quantitative measures of syllabic isochrony, patients with pathology restricted to the cerebellum were asked to produce visually displayed target words embedded into the carrier phrase "Ich habe geC1VC2e gehört" (stop consonants C1 and C2 were selected from the set /p/, /t/, and /k/, C1 and C2 being always identical for a given target; vowel V was chosen from the sample /i/, /u/, /y/, and /a/; examples of target words are *getate*, *gepipe*, etc.). The length of five successive syllables, including the target word, was measured. Compared to controls and subjects suffering from Parkinson's disease, syllable durations were significantly prolonged and there was a tendency toward isochronous spacing in the productions of the patients with cerebellar pathology (for a review, see Hertrich & Ackermann, 1997a).

Phonological vowel-duration contrasts provide a further demonstration of equalized syllable lengths in cerebellar patients. The German language includes word pairs exclusively differing in vowel length—for example, *Rate* (/ra:t^{h}e/, [installment]) versus *Ratte* (/rat^{h}e/, [rat]). In this instance, the distinction between short and long vowels conveys linguistic-phonological information (Hertrich & Ackermann, 1997b). Asked to produce the target words *gepVpe* (V = /a, a:, i, i:, y, y:, u, u:/) embedded into a carrier phrase, cerebellar subjects again showed a proclivity toward isochronous syllable durations insofar as short vowels were lengthened compared to the longer cognates (Ackermann, Hertrich, & Scharf, 1995).

Oral diadochokinesis is evaluated by asking individuals to repeat syllables as fast as possible. Reduced maximum iteration rates under these conditions represent a widely recognized global marker of impaired speech performance. Indeed, syllable repetitions and sentence utterances may represent different types of movements in terms of the underlying control mechanisms. Oral diadochokinesis tasks, nevertheless, provide a rough estimate of maximum speaking rate and, thus, should define an upper limit for speech tempo of sentence productions. A variety of studies reported impaired oral diadochokinesis in subjects suffering from cerebellar disorders or ataxic syndromes (for a review, see Ackermann & Hertrich, 2000). Comparing two different rates of syllable repetitions, Kent, Kent, Rosenbeck, Vorperian, and Weismer (1997) found higher durational irregularity under the fast (as compared to the slow) condition. Control participants showed the opposite pattern. Conceivably, thus, dissolution of syllabic timing in cerebellar disorders evolves across two stages: (a) increased instability of the temporal organization of syllabic sequences, developing finally into (b) slowed and isochronous syllable pacing.

In summary, three aspects of syllable timing were found disrupted in patients with cerebellar lesions: (a) maximum syllable repetition rates during oral diadochokinesis tasks, (b) generation of a normal speaking rate ranging from about 4 Hz to 6 Hz, and (c) the ability to vary syllable length contrasts within sentences.

Scaling of Articulatory Gestures

Conceivably, reduced speaking rate in cerebellar disorders just reflects slowed performance of articulatory gestures. In accordance with these suggestions, preliminary kinematic analyses provided direct evidence of prolonged movement times as well as reduced peak velocity of orofacial excursions in ataxic dysarthria (for a review, see Ackermann & Hertrich, 2000). Isolated kinematic measures, however, do not provide an adequate description of speech motor performance. Rather, the ratio of maximum velocity and displacement is considered an important control parameter of skilled speech motor processes (e.g., Kelso, Vatikiotis-Bateson, Saltzman, & Kay, 1985). To systematically evaluate the relationship of peak velocity and maximum range in cerebellar patients, lip excursions were registered by means of a three-dimensional optoelectric movement analysis system during the production of the test sentences “Ich habe gepape gelesen” or “Ich habe gepappe gelesen” (Ackermann et al., 1995, Ackermann, Hertrich, Daum, Scharf, & Spieker, 1997). Because short and long vowels represent distinct phonemes in German, the target words were produced with either short or long vowels. Control participants and patients with cerebellar lesions showed a highly linear relationship between peak velocity and movement range. In contrast, the patients exhibited decreased velocity-displacement ratios—that is, slower lip excursions at given amplitudes. The deficit varied according to the type of movement and, to a lesser degree, the linguistic demands of the task. Between-group differences were found amplified during closing as compared to opening gestures and during short versus long vowels. These data point at an impaired ability to generate adequate muscular forces under temporal constraints. Cerebellar dysarthria, thus, is associated with compromised execution of single vocal-tract gestures.

Articulatory Sequencing

Besides force scaling of single articulatory gestures, speech production requires the adequate temporal coordination (sequencing) of respiratory, laryngeal, and orofacial muscle activities. For example, the release of vocal-tract occlusion and the subsequent initiation of glottal vibrations, respectively, have to be properly adjusted during the production of stop consonant-vowel sequences (see Figure 1A for an example; see explanation of Figure 1 in Cerebellar Contributions to Speech Perception: Clinical and Functional-Imaging Data section). Thus, the time lag between stop consonant burst and vowel onset—voice onset time (VOT)—can be considered a measure of the timing of orofacial and laryngeal events (Keller, 1990). Voiced and unvoiced stop consonants show a qualitative or categorical distinction of VOT (e.g., Lisker & Abramson, 1964, 1967); the former are characterized by shorter VOT values (short-lag VOT) than their unvoiced cognates (long-lag VOT) or even present with a voicing lead—that is, glottal vibrations starting prior to stop consonant release burst. At the initial position of isolated English words, VOT of, for example, the voiced stop /d/ falls below about 20 ms, whereas this parameter in case of unvoiced /t/ exceeds 30 ms. Similar data have been obtained from German speakers (see Ackermann & Hertrich, 1997). Thus, VOT conveys phonological information and this durational parameter must be expected to be explicitly controlled. Ivry and Gopal (1992) found subjects with cerebellar atrophy to display greater VOT variability during production of (voiced and unvoiced) stop consonant-vowel syllable pairs such as /ba/ and /pa/. A subsequent study asked patients with an ataxic syndrome resulting from cerebellar degeneration to produce sentence utterances including either one of the German minimal pair cognates *Daten* and *Taten* (Ackermann & Hertrich, 1997). In line with the preceding investigation, the cerebellar subjects again showed a reduced categorical distinction of the VOT of voiced and unvoiced stop consonants. Most noteworthy, disruption of long-lag VOT primarily accounted for these findings. Because unvoiced stops pose higher demands on the coordination of vocal-tract muscles than their unvoiced cognates (Cooper, 1977), the findings of a predominant disorder of long-lag VOT support the assumption of cerebellar participation in the sequencing of orofacial and laryngeal activities during speech production.

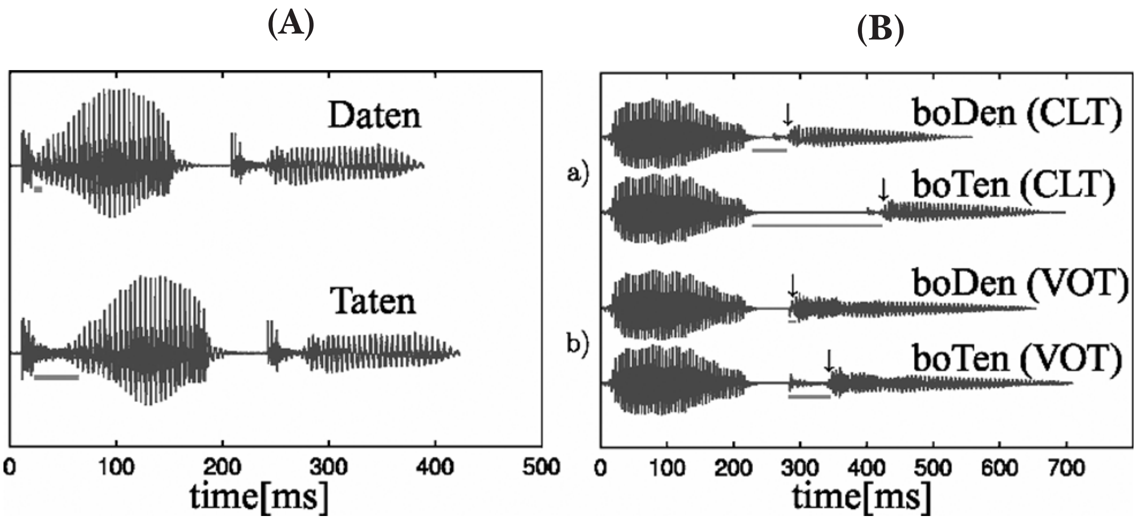

Figure 1: (A) Oscillogram (i.e., the trajectory of sound-pressure level across time) of the Utterances *Taten* ([t^{h}atn], *deeds*) and *Daten* ([datn], *data*) as Produced by a Normal Speaker and (B) Prototypical Examples of the Two Series of *Boden–Boten* Utterances.

NOTE: (A) The horizontal lines indicate voice onset time (VOT)—that is, the time lag between stop consonant burst and the onset of phonation (in this case vowel /a/). VOT is considered a measure of the temporal coordination of the underlying successive articulatory gestures. Unvoiced stop consonants, such as the word-initial /t/ in *Taten*, are characterized by longer VOT (> 30 ms) than their voiced cognates, such as /d/ in *Daten* (< 20 ms). As a prerequisite of their perceptual differentiation, VOT of voiced and unvoiced stop consonants do not overlap at a word-initial position. (B) The first two waveforms differ in closure time (CLT)—that is, the underlined pause signaling word-medial /d/ or /t/, respectively, all the other components being identical (arrows = voice onset of the second syllable). The utterance with short CLT ([bodn]) is perceived as the word *Boden* and the one with long CLT ([botn]) as *Boten*. The lower two acoustic speech signals differ only inVOT—that is, the underlined noise segment extending from the burst to the beginning of the vocalic segment (arrows = onset of the schwa vowel). A series of equally spaced stimuli was derived from these two pairs of acoustic signals (CLT: 27 ms to 167 ms; VOT: 9 ms to 69 ms). The arrows mark the onset of the voiced portion of the second syllable.

FUNCTIONAL MAGNETIC RESONANCE IMAGING (fMRI) INVESTIGATIONS OF CEREBELLAR INVOLVEMENT IN OVERT AND COVERT SPEECH

To some extent, the VOT abnormalities observed in cerebellar disorders resemble those found in apraxia of speech (see Ackermann & Hertrich, 1997). The latter syndrome is considered a higher order dysfunction of speech motor control characterized by impaired "programming" of articulatory gestures (e.g., Dronkers, 1996). Motor imagery has been assumed to provide a window into the mechanisms of action planning (Jeannerod, 1994). Although such tasks have been widely used for investigating the neural bases of manual actions, less attention has been given to covert speech. Functional imaging of internal speech, thus, should allow for the delineation of the cerebral network subserving prearticulatory motor control processes for verbal utterances. In two fMRI studies (Ackermann, Wildgruber, Daum, & Grodd, 1998; Riecker, Ackermann, Wildgruber, Dogil, & Grodd, 2000), participants were asked to repeat, either overtly or covertly, highly overlearned word strings, such as the names of the months of the year. This simple task ("automatic speech") was considered a rather selective probe of speech motor control, given the relatively low demands placed on syntactic and semantic processing. Whereas other investigations of the motor aspects of speech production asked subjects to repeat or read single nouns (Petersen, Fox, Posner, Mintun, & Raichle, 1989; Wise, Greene, Büchel, & Scott, 1999), automatic speech provides an opportunity to address fluent or connected verbal output. These test materials must be expected to elicit a rather monotonous mode of speech production that should focus on segmental aspects of verbal utterances and, thus, minimize confounding influences of intonational (suprasegmental) patterns.

Participants were asked to perform these tasks at their habitual speaking rate. Analysis of the group data revealed exclusive activation of the left precentral gyrus and right cerebellar hemisphere during covert performance (task versus resting baseline; see Figure 2). The supratentorial cortical activation emerged along the lower extent of precentral gyrus (Brodmann area 6).

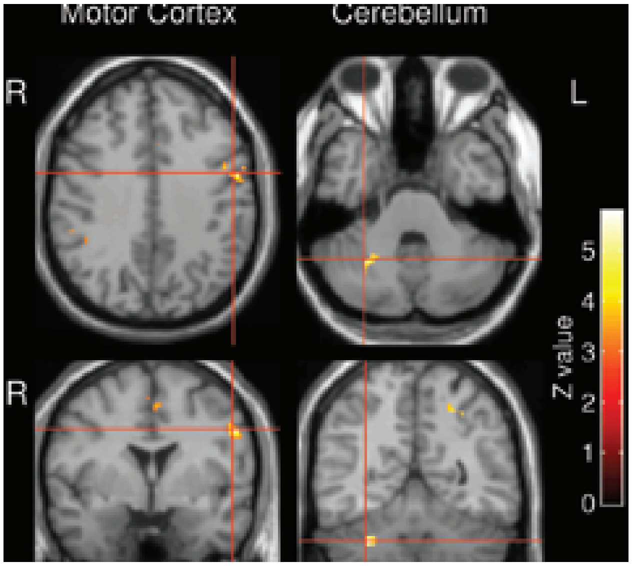

Figure 2: Functional Magnetic Resonance Imaging (fMRI) Activation Maps ($Z > 3.7$) Obtained During Silent Production of "Automatic Speech" (i.e., continuous recitation of the months of the year) Superimposed on the Averaged T1-Weighted Anatomical Images Across 18 Subjects in TALAIRACH Space at the Level of Motor Cortex (left column) and the cerebellum (right column; upper row = transverse slices, lower row = coronal slices).

Thus, the cerebellum even participates in the generation of "internal speech." Overt task performance yielded rather bilateral hemodynamic activation of these areas, concomitant with moderate lateralization effects toward the same direction as in the imagery conditions. Conceivably, the broader distribution of hemodynamic responses in association with overt speech reflects the additional computational demands associated with overt speech production, such as force scaling (see above).

As discussed previously, slowed syllable rate during oral diadochokinesis tasks or sentence utterances represents a salient feature of cerebellar speech disorders. However, speech tempo, as a rule, does not fall below about 3 Hz (Ackermann et al., 1995; Hertrich & Ackermann, 1997a). Therefore, cerebellar activation across a series of syllable repetition rates might be expected to be limited to rates higher than 3 Hz. To test this hypothesis, neurologically healthy subjects were asked to produce repetitions of the syllable /ta/ at three different rates (2.5 Hz, 4.0 Hz, and 5.5 Hz) during fMRI measurements, performance being paced via earphones by recorded trains of the syllable /ta/ at the respective repetition frequencies (Wildgruber, Ackermann, & Grodd, 2001). Syllable iterations were carried out in a covert or inner speech mode. Indeed, as compared to baseline, significant fMRI effects at the level of the cerebellum were found only at repetition rates of 4.0 Hz and 5.5 Hz. The amplitude of hemodynamic responses at local activation maxima within the motor cortex showed a positive linear relationship to syllable production rate. Cerebellar activation, by contrast, exhibited a clear-cut threshold phenomenon in terms of almost no signal increase during 2.5 Hz, concomitant with considerably enlarged but similar magnitudes of the responses bound to 4.0 Hz and 5.5 Hz syllable repetitions (see Figure 3).

CEREBELLAR CONTRIBUTIONS TO SPEECH PERCEPTION: CLINICAL AND FUNCTIONAL-IMAGING DATA

The first evidence of a perceptual deficit associated with cerebellar pathology was observed on a task in which the participants judged the duration of time intervals of around 400 ms (Ivry & Keele, 1989). Coupled with work on motor timing and motor learning, these findings corroborated the suggestion (Braitenberg, 1967) that the cerebellum can be conceptualized as an internal timing system (for a review, see Ivry, 1997). The latter hypothesis has led to investigations of cerebellar involvement in speech perception for sound contrasts that require precise timing. Both the English and the German languages include pairs of syllables or lexical items exclusively differing in a single durational parameter of the acoustic signal (see Figure 1b). Assuming an essential contribution of the cerebellum to time perception, distorted recognition of temporally cued speech sounds might be expected in patients with cerebellar lesions. The first studies (Ackermann, Gräber, Hertrich, & Daum, 1997; Ivry & Gopal, 1992) failed to detect any impairments when cerebellar patients were required to categorize stop consonant-vowel syllables varying in VOT. However, these stimuli differ not only in durational aspects but also in their spectral content (see Ackermann, Gräber, et al., 1997; Ackermann, Gräber, Hertrich, & Daum, 1999). To establish a situation in which the only cue of a speech sound contrast lies within the temporal domain, Ackermann and coworkers (Ackermann, Gräber, et al., 1997) created a stimulus continuum in which they varied the length of a silent period created by the occlusion preceding a medial stop consonant. For the word *Boden* (bodn), this pause is short while the lexical item *Boten* (botn) encompasses a longer occlusion time (again, see Figure 1B). Patients with diffuse cerebellar atrophy were unable to discriminate sounds along this continuum.

Cerebellar contribution to the perceptual analysis of speech sound contrasts defined by temporal cues is further supported by a recent fMRI study (Mathiak,

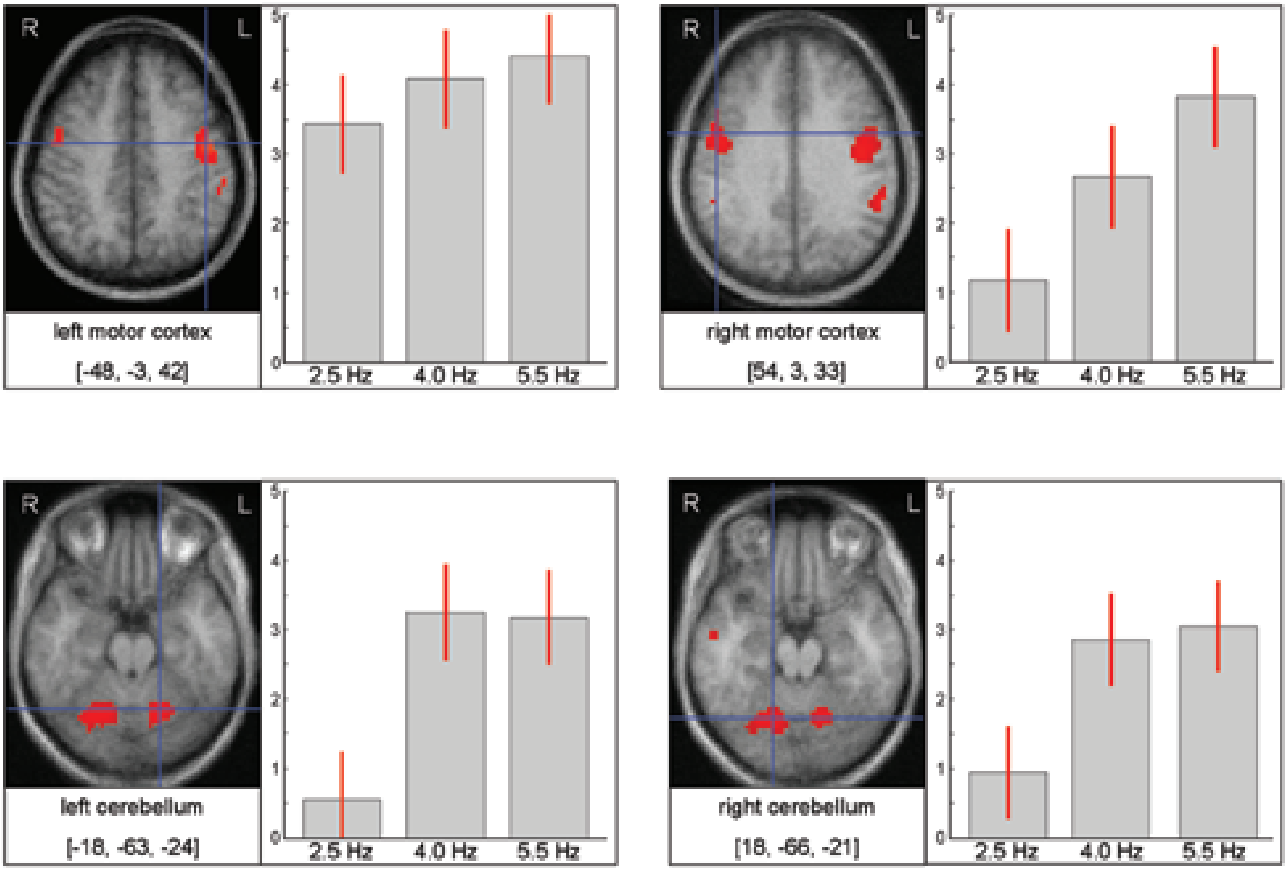

Figure 3: Parametric Analysis of Hemodynamic Activation During Covert Syllable Repetitions (auditory imagery).
NOTE: The TALAIRACH coordinates (in square brackets) refer to the highest activated voxel within each cluster displayed on transversal slices of the averaged anatomical reference images. Size and variance of signal intensity within the respective voxel (in arbitrary units) is given for all three production frequencies considered (Statistical Parametric Mapping (SPM99), $n = 10$, $T > 3.09$, $p < .001$ uncorrected). Note the threshold effect at the level of the cerebellum.

Hertrich, Grodd, & Ackermann, 2002). In this experiment, participants again had to judge whether a stimulus sounds like *Boden* or like *Boten*. In contrast to the preceding clinical investigation, two different continua of acoustic signals were considered as test materials (as shown in Figure 1B). Similar to the previous patient study, the first series of stimuli varied the silent period associated with word-medial occlusion (duration ranged from 27 ms to 167 ms). In these instances, categorization of the applied items solely depends on the analysis of a temporal parameter—that is, occlusion duration. Concerning the second continuum, the duration of word-medial occlusion was held constant at 70 ms, and the distinction between the /d/ and /t/ sound was created by varying the VOT associated with the release of the stop consonant. Because the period of aspirated noise varies under these conditions, discrimination can rely on temporal as well as spectral parameters of the acoustic speech signal. Both types of cues can be found in normal speech. The pure temporal continuum corresponds to speech during normal conversation in which the syllables tend to become blended together, whereas variation in VOT (and thus the introduction of spectral cues) occurs in more formal, well-articulated speech.

Although it is possible to perceive differences between the two stimulus sets, subjects' reaction times did not significantly vary nor was there any difference in accuracy. Thus, the two tasks were of similar difficulty and, phenomonologically, seemed identical. Nonetheless, activation within the cerebellum was found significantly enlarged when subjects listened to the occlusion series compared to the VOT series. This hemodynamic response was restricted to a single cluster of voxels rostral to the horizontal fissure within the right hemisphere (see Figure 4; TALAIRACH coordinates of highest *z* score [20-82-16]). The only other area being more active in the occlusion condition was a region anterior and inferior to Broca's area. In contrast, the reversed subtraction (VOT minus occlusion condition) yielded an exclusive hemdynamic response within the left

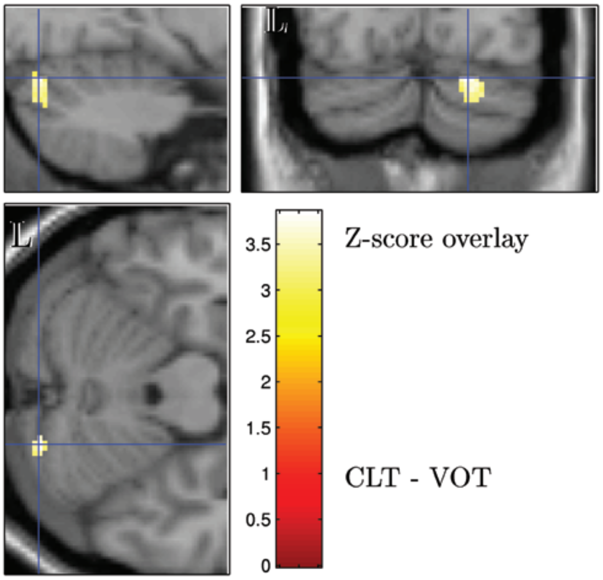

Figure 4: Significant BOLD Effects at the Level of Infratentorial Structures (z score overlay on normalized anatomical images, threshold at $z > 3.1$, corresponds to $p < .001$, uncorrected).

NOTE: Within the right cerebellar hemisphere, decoding of the intersegmental closure time (CLT) yielded a significantly higher BOLD effect than the processing of intrasyllabic voice onset time (VOT). The displayed activation cluster pertains to lateral aspects of Crus I of the neocerebellum (see Schmahmann, Doyon, Toga, Petrides, & Evans, 2000).

supratemporal plane. It is interesting to note that the paravermal superior right cerebellar response overlaps with cerebellar activation sites in the studies of covert and overt speech (e.g., Wildgruber et al., 2001). This area, thus, might subserve temporal computations within the domains of both expressive and perceptual speech functions.

Liberman and Mattingly (1989, reprinted in Liberman, 1996) introduced the notion of a "phonetic module" to explain a variety of psychophysical data on speech processing, such as duplex perception. Modular input systems are basically characterized by "informational encapsulation" and a "distinct neural architecture" (Fodor, 1996). The phonetic module has been associated with left-hemisphere "language" areas (Liberman, 1996; Mathiak, Hertrich, Lutzenberger, & Ackermann, 2001). More recently, it has been proposed that speech encoding involves bilateral posterior superior temporal cortex (Poeppel, 2001). The findings reviewed here suggest that ideas about language modularity—in particular, within the domains of speech perception and production—need to be reconsidered. One reconceptualization would be to broaden the range of neural areas that are pertaining to this module, encompassing the cerebellum as part of the respective cerebral network. Alternatively, it may be more parsimonious to consider language as recruiting a series of modules that perform specialized operations that do not need to be language specific. From this perspective, the cerebellum is hypothesized to be engaged when speech tasks require precise temporal representations.

MODEL: CEREBELLAR TRANSFORMATIONS ON SYLLABLE-SIZED ARTICULATORY PROGRAMS RETRIEVED FROM A PREMOTOR MENTAL SYLLABARY

The phonotactic rules of English allow for the generation of more than 12,000 different syllables (Levelt, 2001; Levelt, Roelofs, & Meyer, 1999). However, the 500 highest ranked items suffice to handle about 80% of verbal utterances. These syllables, by definition, involve the most exercised motor activities and, thus, can be considered highly overlearned movement patterns. Storage of at least the highest ranked syllables as preprogrammed motor routines should considerably reduce the computational demands during speech production as compared to online assembly of single phonetic features or the respective orofacial-laryngeal excursions for the production of less frequent speech sounds. Indeed, the initiation of high-frequency syllables is faster than for low-frequency cognates (Levelt & Wheeldon, 1994). Levelt and coworkers proposed a mental syllabary associated with premotor cortex housing syllable-sized, overlearned articulatory programs. By contrast, word forms comprising two or more syllables are not stored as prespecified motor routines. Consequently, sequencing of syllables into larger utterances must be performed online during speech production and, thus, should depend on the motor execution system.

The model of a mental syllabary assumes that the stored canonical gestural programs represent coarticulated units (Levelt et al., 1999). However, coarticulatory influences extend across syllable boundaries (see Hertrich & Ackermann, 1999), an effect that is attributed to the motor execution system rather than the mental syllabary (Levelt et al., 1999). Such cross-syllable coarticulation effects are diminished in patients with cerebellar dysarthria, even though these patients did exhibit normal temporal modulations within a syllable (Hertrich & Ackermann, 1999).

Modulation of syllable lengths represents an information-bearing aspect of speech production, contributing to the meaning of the utterance as well as helping convey the speaker's emotional state. Given that coarticulation effects extend across syllables, stored canonical articulatory routines cannot be simply concatenated together. Similarly, to increase speaking rate, it is not sufficient to shorten interword and intersyllable pauses. Rather, the length of the retrieved motor programs must be adjusted according to articulatory, linguistic, and emotional constraints. Considering the dissolution of the temporal organization of verbal utterances in ataxic dysarthria, a process evolving into slowed and isochronous pacing, the cerebellum represents a candidate structure for the operations required to establish speech tempo and the suprasegmental shaping of syllable templates retrieved from a frontal store.

Based on morphological and electrophysiological data, Braitenberg, Heck, and Sutlan (1997) proposed that the cerebellar cortex acts as a "sequence-in/sequence-out operator" transforming input sequences of events into a coordinated output sequence. Applying this model to the domain of speech production, the cerebellum "sculpts" rhythmically structured verbal utterances at a speaker's habitual speech tempo of 4 to 6 syllables per second out of a string of syllable templates. We propose that a left-fronto/right-cerebellar network subserves the ongoing sequencing of a prearticulatory speech code that, in turn, steers the subsequent coordination of vocal-tract muscle innervation.

In addition to the assumed disordered temporal organization of syllable strings at a prearticulatory level, cerebellar patients are impaired in speech-perception tasks (word recognition) requiring the encoding of durational parameters of the acoustic signal. The hemodynamic responses associated with these two aspects of verbal-acoustic communication—that is, internal speech and speech perception—were found to be organized along the rostro-caudal direction within superior paravermal aspects of the right cerebellar hemisphere. As a consequence, this area seems to provide a common platform for the computation of temporal parameters of the sound structure of human languages across both output and input mechanisms. Most noteworthy, the auditory word identification task found impaired in cerebellar patients (*Boden* versus *Boten*) requires calculation of an intersyllabic durational parameter—that is, occlusion time (see above). By contrast, estimation of an intrasyllabic length measure (VOT) was uncompromised. Perceptual encoding of occlusion time, thus, also might depend on cerebellar representation of precise syllable sequences at the level of the prearticulatory speech code.

IMPLICATIONS OF THE SPEECH DATA: MECHANISMS OF CEREBELLAR MODULATION OF COGNITIVE FUNCTIONS

Patients with cerebellar lesions frequently show cognitive deteriorations similar to those observed in patients with frontal lobe lesions, although the extent of the impairments may not be as marked (for a review, see Ackermann & Daum, 2003; Ivry & Fiez, 2000). The reciprocal connections between prefrontal cortex and cerebellum have been proposed as the anatomic basis of these disorders (Middleton & Strick, 1994; Schmahmann & Pandya, 1997). Complex cognitive operations, such as that involved in solving mathematical equations or planning extended behaviors, place considerable demands on working memory and are likely to depend on or to be enhanced by "internal speech" processes. The impaired generation of a prearticulatory speech code subsequent to cerebellar dysfunction would disrupt such cognitive operations.

CONCLUSION

Clinical and functional-imaging data indicate that the cerebellum provides a common platform for the computation of at least some temporal parameters of verbal utterances both during speech production and perception. Furthermore, superior paravermal aspects of the right cerebellar hemisphere seem to support the sequencing of syllabic units even at the level of "inner speech" or "auditory verbal imagery." As a consequence, cerebellar disorders can be expected to compromise cognitive operations encompassing the generation of a prearticulatory speech code.

REFERENCES

Ackermann, H., & Daum, I. (2003). Neuropsychological deficits in cerebellar syndromes. In M. A. Bédard, Y. Agid, S. Chouinard, S. Fahn, A. D. Korczyn, & P. Lespérance (Eds.), *Mental and behavioral dysfunction in movement disorders* (pp. 147–156). Totowa, NJ: Humana.

Ackermann, H., Gräber, S., Hertrich, I., & Daum, I. (1997). Categorical speech perception in cerebellar disorders. *Brain and Language*, *60*, 323-331.

Ackermann, H., Gräber, S., Hertrich, I., & Daum, I. (1999). Cerebellar contributions to the perception of temporal cues within the speech and nonspeech domain. *Brain and Language*, *67*, 228-241.

Ackermann, H., & Hertrich, I. (1997). Voice onset time in ataxic dysarthria. *Brain and Language*, *56*, 321-333.

Ackermann, H., & Hertrich, I. (2000). The contribution of the cerebellum to speech processing. *Journal of Neurolinguistics*, *13*, 95-116.

Ackermann, H., Hertrich, I., Daum, I., Scharf, G., & Spieker, S. (1997). Kinematic analysis of articulatory movements in central motor disorders. *Movement Disorders*, *12*, 1019-1027.

Ackermann, H., Hertrich, I., & Scharf, G. (1995). Kinematic analysis of lower lip movements in ataxic dysarthria. *Journal of Speech and Hearing Research*, *38*, 1252-1259.

Ackermann, H., Wildgruber, D., Daum, I., & Grodd, W. (1998). Does the cerebellum contribute to cognitive aspects of speech production? A functional magnetic resonance imaging (fMRI) study in humans. *Neuroscience Letters*, *247*, 187-190.

Akshoomoff, N. A., Courchesne, E., & Townsend, J. (1997). Attention coordination and anticipatory control. In J. D. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 575-598). San Diego, CA: Academic Press.

Baddeley, A. D. (1992). Working memory: The interface between memory and cognition. *Journal of Cognitive Neuroscience*, *4*, 281-288.

Braitenberg, V. (1967). Is the cerebellar cortex a biological clock in the millisecond range? *Progress in Brain Research*, *25*, 334-346.

Braitenberg, V., Heck, D., & Sultan, F. (1997). The detection and generation of sequences as a key to cerebellar function: Experiments and theory. *Behavioral and Brain Sciences*, *20*, 229-277.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, *12*, 1-47.

Chein, J. M., & Fiez, J. A. (2001). Dissociation of verbal working memory system components using a delayed serial recall task. *Cerebral Cortex*, *11*,1003-1014.

Cooper, W. E. (1977). The development of speech timing. In S. J. Segalowitz & F. A. Gruber (Eds.), *Language development and neurological theory* (pp. 357-373). New York: Academic Press.

Dolan, R. J. (1998). A cognitive affective role for the cerebellum. *Brain*, *121*, 545-546.

Dow, R. S., & Moruzzi, G. (1958). *The physiology and pathology of the cerebellum*. Minneapolis: University of Minnesota Press.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, *384*, 159-161.

Fiez, J. A., Raife, E. A., Balota, D. A., Schwarz, J. P., Raichle, M. E., & Petersen, S. E. (1996). A positron emission tomography study of the short-term maintenance of verbal information. *Journal of Neuroscience*, *16*, 808-822.

Fodor, J. A. (1996). *The modularity of mind: An essay on faculty psychology* (10th ed.). Cambridge, MA: MIT Press.

Frackowiak, R. S. J., Friston, K. J., Frith, C. D., Dolan, R. J., & Mazziotta, J. C. (1997) *Human brain function*. San Diego, CA: Academic Press.

Hallett, M., & Grafman, J. (1997). Executive function and motor skill learning. In J. D. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 297-323). San Diego, CA: Academic Press.

Hertrich, H., & Ackermann, H. (1997a). Acoustic analysis of durational speech parameters in neurological dysarthrias. In Y. Lebrun (Ed.), *From the brain to the mouth: Acquired dysarthria and dysfluency in adults* (pp. 11-47). Dordrecht, the Netherlands: Kluwer.

Hertrich, I., & Ackermann, H. (1997b). Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures. *Journal of the Acoustical Society of America*, *102*, 523-536.

Hertrich, I., & Ackermann, H. (1999). Temporal and spectral aspects of coarticulation in ataxic dysarthria: An acoustic analysis. *Journal of Speech, Language, and Hearing Research*, *42*, 367-381.

Ivry, R. (1997). Cerebellar timing systems. In J. D. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 555-573). San Diego, CA: Academic Press.

Ivry, R. B., & Fiez, J. A. (2000). Cerebellar contributions to cognition and imagery. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 999-1011). Cambridge, MA: MIT Press.

Ivry, R. B., & Gopal, H. S. (1992). Speech production and perception in patients with cerebellar lesions. In D. E. Meyer & S. Kornblum (Eds.), *Synergies in experimental psychology, artificial intelligence and cognitive neuroscience* (pp. 771-802). Hillsdale, NJ: Lawrence Erlbaum.

Ivry, R. B., & Keele, S. W. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience*, *1*, 136-152.

Ivry, R. B., Justus, T. C., & Middleton, C. (2001). The cerebellum, timing, and language: Implications for the study of dyslexia. In M. Wolf (Ed.), *Dyslexia, fluency, and the brain* (pp. 189-211). Timonium, MD: York Press.

Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*, 187-245.

Keller, E. (1990). Speech motor timing. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 343-364). Dordrecht, the Netherlands: Kluwer.

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

Kent, R., & Rosenbek, J. (1982). Prosodic disturbance and neurologic lesion. *Brain and Language*, *15*, 259-291.

Kent, R. D., Kent, J. F., Rosenbek, J. C., Vorperian, H. K., & Weismer, G. (1997). A speaking task analysis of the dysarthria in cerebellar disease. *Folia Phoniatrica & Logopaedica*, *49*, 63-82.

Lechtenberg, R. (1993). *Handbook of cerebellar diseases*. New York: Marcel Dekker.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, *98*, 13464-13471.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-75.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 239-269.

Liberman, A. M. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.

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

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

Mathiak, K., Hertrich, I., Grodd, W., & Ackermann, H. (2002). Cerebellum and speech perception: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *14*, 902-912.

Mathiak, K., Hertrich, I., Lutzenberger, W., & Ackermann, H. (2001). Neural correlates of duplex perception: A whole-head magnetencephalography study. *NeuroReport*, *12*, 501-506.

Middleton, F. A., & Strick, P. L. (1994). Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science*, *266*, 458-461.

Molinari, M., Leggio, M. G., & Silveri, M. C. (1997). Verbal fluency and agrammatism. In J. D. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 325-339). San Diego, CA: Academic Press.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintum, M., & Raichle, M. E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience*, *1*, 153-170.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, *25*, 679-693.

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., & Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *NeuroReport*, *11*, 1997-2000.

Riecker, A., Wildgruber, D., Dogil, G., Grodd, W., & Ackermann, H. (2002). Hemispheric lateralization effects of rhythm implementation during syllable repetitions: An fMRI study. *NeuroImage*, *16*, 169-176.

Schmahmann, J. D., Doyon, J., Toga, A. W., Petrides, M., & Evans, A.C. (2000). *MRI atlas of the human cerebellum*. San Diego, CA: Academic Press.

Schmahmann, J. D., & Pandya, D. N. (1997). The cerebrocerebellar system. In J. D. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 31-60). San Diego, CA: Academic Press.

Schmahmann, J. D., & Sherman, J. C. (1998). The cerebellar cognitive affective syndrome. *Brain*, *121*, 561-579.

Wildgruber, D., Ackermann, H., & Grodd, W. (2001). Differential contributions of motor cortex, basal ganglia, and cerebellum to speech motor control: Effects of syllable repetition rate evaluated by fMRI. *NeuroImage*, *13*, 101–109.

Wise, R. S. J., Greene, J., Büchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *Lancet*, *353*, 1057-1061.