Timing speech: a review of lesion and neuroimaging findings

Annett Schirmer*

Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany

Accepted 15 April 2004
Available online 19 June 2004

Abstract

Time is a fundamental dimension of behavior and as such underlies the perception and production of speech. This paper reviews patient and neuroimaging studies that investigated brain structures that support temporal aspects of speech. The left-frontal cortex, the basal ganglia, and the cerebellum represent structures that have been implicated repeatedly. A comparison with the structures involved in the timing of non-speech events (e.g., tones, lights, finger movements) suggests both commonalities and differences: while the basal ganglia and the cerebellum contribute to the timing of speech and non-speech events, the contribution of left-frontal cortex seems to be specific to speech or rapidly changing acoustic information. Motivated by these commonalities and differences, this paper presents assumptions about the function of basal ganglia, cerebellum, and cortex in the timing of speech.

© 2004 Elsevier B.V. All rights reserved.

Theme: Neural basis of behavior
Topic: Cognition

Keywords: Prosody; Duration; Tempo; VOT; CLT; Lateralization; Basal ganglia; Cerebellum

1. Introduction

Spoken language comprises various sounds that rapidly change in time. To ensure accurate speech comprehension, speakers must time the production of individual sounds precisely. This necessity arises because speech perception and comprehension are largely determined by temporal order and duration. The specific ways in which temporal information determines lexical and syntactic aspects of speech vary across languages. For example, some languages such as Japanese distinguish between short (e.g., /isso/ rather) and long sounds (e.g., /isso:/ more), whereas other languages such as Estonian use three classes of sound lengths: short, long, and overlong. Furthermore, languages might differ with respect to the acoustic information that accompanies modulations in sound length. In English and German, for example, long vowels are not simply longer than short vowels but are also produced with heightened subglottal air pressure. Thus, long vowels are referred to as tense and contrasted with short, lax vowels [37].

However, even in languages that show redundancies in the coding of linguistic meaning, listeners are trained to exploit temporal information. Evidence supporting this conclusion comes from experiments that investigate speech comprehension under conditions of reduced spectral information (e.g., Ref. [109]). Replacing much of the spectral information (i.e., frequency bands) in words and sentences with noise while leaving temporal information preserved seems to have only little effect on speech comprehension.

1.1. The temporal information embedded in speech

The information that can be derived from duration guides a variety of linguistic processes ranging from the perception of individual speech segments (e.g., consonants) to the processing of syntactic ambiguities. The following paragraphs introduce the phenomena that have been studied most frequently with respect to their neuroanatomical underpinnings. As the great majority of these studies are based on Asian languages such as Thai and Indo-European languages such as English and German, the discussion will be restricted to the temporal phenomena that are relevant in these languages.

E-mail address: schirmer@cbs.mpg.de (A. Schirmer).
At the level of individual speech segments, duration is critical for the perception of stop consonants such as ‘d’ or ‘t’. As illustrated in Fig. 1, the percept ‘d’ as in ‘ladder’ is elicited when the interval between the initial burst of the consonant and the onset of voicing (voice onset time, VOT) is shorter than 30 ms. In contrast, longer VOTs will elicit the percept ‘t’ as in ‘latter’ [71]. A small period of silence (closure time, CLT) that precedes stop consonants in a word’s medial position has similar importance. Listeners perceive the word ‘ladder’ correctly with CLTs smaller than 40 ms, while longer CLTs will change the percept to ‘latter’ [68].

In addition to the significance of duration for the distinction among stop consonants, time plays a role at a higher-order level that subsumes individual speech segments to form syllables, words, and sentences. The acoustic properties of this higher-order level are also referred to as prosody and can be described as modulations in speech melody (fundamental frequency, F0), amplitude, and duration. Speakers modulate these parameters in order to express linguistic and emotional meaning. Stress is an example for the linguistic function of prosody. Stressed syllables have a higher F0, are louder and longer than unstressed syllables. Furthermore, prosody can convey information concerning the syntactic structure of an utterance. To that effect, words at the end of syntactic boundaries are often lengthened, followed by a short pause, and marked by a specific boundary tone. These cues can predict the syntactic role of an up-coming sentence constituent. That ongoing sentence processing is disrupted when the syntactic role of a constituent does not meet prosodically induced expectations [110,120] strongly suggests that listeners rely on these cues during speech comprehension.

In addition to lexical and syntactic information, prosody also reflects a speaker’s emotional state. The tight coupling of the vocal expression system with physiological parameters such as blood pressure and respiration makes speech very susceptible to emotions [105]. Moreover, the physiological changes that accompany different emotions have a specific effect on speech that allows the identification of a speaker’s emotional state independent of his or her cultural background [104]. For example, an utterance spoken in a happy emotional state is usually faster than the same utterance spoken in a sad emotional state, as well as being louder and having a broader F0 range. In this context, speech tempo correlates with the arousal of the speaker.

1.2. Neuroanatomical correlates of language and prosody

The question of how listeners encode the linguistic and emotional message that is conveyed by language and its acoustic wrapping called prosody and what brain structures mediate this process has been of interest to researchers for several decades (e.g., Refs. [59,83]). Investigations have concerned both the processes that underlie written and spoken language comprehension and the processes that extract linguistic and emotional meaning from prosody. With respect to the processes involved in language, lesion and neuroimaging studies highlight the significance of left frontal and temporal structures [29,66,113,114]. Although both hemispheres seem to contribute to language processing, most researchers agree upon a special role of the left hemisphere especially with respect to phonological and syntactic processes that follow the early construction of sound-based representations of speech [35,56,61,108]. Interestingly, this is also true for language systems that are based on signs [14].

In contrast to language comprehension, lateralization of prosodic processes is less clear. Observations of compromised prosodic functions have been reported following both right and left hemisphere damage. While some findings led to suggest that prosody is functionally lateralized—the right hemisphere being specialized for emotional prosody and the left hemisphere being specialized for linguistic prosody [93,116,117], evidence of both linguistic and emotional prosodic deficits following insult to the right hemisphere challenged this proposal [19,121]. A recent review of the literature [36] suggests that a different hypothesis might better account for the currently available evidence. According to this proposal, the degree to which prosodic features are segmentally bound determines lateralization [123]. Prosodic expression that extends over a longer phrase or sentence (e.g., intonation) is thought to be lateralized to the right hemisphere, whereas prosodic expression that is associated with a syllable (e.g., tone) is thought to be ‘drawn’ toward the left hemisphere.

Note that besides the neocortex, subcortical structures such as the striatum have been implicated in prosody, thus indicating the importance of cortico-striatal circuits [20,22,34,119]. The implication of these structures is especially important to the current discussion, as they are also involved in the timing of non-speech events [53,78,79,86].

Most of the aforementioned proposals of speech and prosody are based on a holistic research approach; speech and prosody were investigated as a whole. However, an analytic strategy that examines the processing of their spectral and temporal components might reveal a more

---

Fig. 1. The words ‘ladder’ and ‘latter’ differ in CLT (light gray box) and in VOT (dark gray box). Both CLT and VOT are shorter for the voiced as compared to the unvoiced stop consonant.
detailed picture. This latter approach has already been adopted by several researchers. The present paper provides an overview of the work that has been related to the neuroanatomical correlates of processing time in speech and prosody. To this end, the paper is divided into different sections that address findings from patients with cortical lesions, with basal ganglia dysfunction, and with cerebellar disorders. Furthermore, neuroimaging findings from healthy participants will be presented and the implications from both patient and neuroimaging studies will be related to research that investigated the timing of non-speech events.

2. Patients with cortical lesions

Given that duration is such an important dimension of speech, one might assume that the language dominant hemisphere processes speech inherent temporal information. Evidence supporting this assumption comes from clinical observations in patients with Broca’s aphasia—a neurological disorder that frequently follows insult to frontal regions in the left hemisphere. The speech of Broca’s aphasics can be characterized as agrammatic, nonfluent, and dysprosodic [48]. Moreover, these patients exhibit a slower than normal speaking rate [99]. This deficit is neuroanatomically specific as lesions in the more posterior part of the dominant hemisphere often result in a different form of aphasia called Wernicke’s or fluent aphasia. These patients speak with a relatively natural language rhythm. However, Wernicke’s aphasics frequently show errors of word and phoneme selection so that their speech is semantically empty [48,51]. Although a differentiation between the forms of aphasia is useful, experimental investigations of temporal processing following cortical insult frequently cluster all aphasics into one group and compare their performance to that of healthy controls and patients with right hemisphere lesions. Consequently, the following presentation of studies can refer to the specific forms of aphasia and the different lesion sites only if the studies allow such a distinction.

Early experimental evidence for a temporal deficit in aphasia was reported by Efron [31]. He asked patients with left or right hemisphere lesions to indicate which of two tones (10 ms duration) or light pulses (5 ms duration) occurred first. Intervals between two events ranged from 0 to 600 ms. Efron [31] found that patients with deficient temporal sequencing had a lesion in the language dominant hemisphere and exhibited some degree of aphasia. Furthermore, he noticed that although these patients demonstrated impaired temporal sequencing for both auditory and visual stimuli, the impairment was more severe in the auditory modality. Based on these findings, Efron [31] concluded that aphasia reflects a primary defect in temporal analysis. Moreover, he proposed that left hemisphere damage (LHD) disrupts the processes that time-label each incoming speech sound so that words like ‘pots’ are scrambled and form words like ‘tops’ or non-existing strings like ‘sopt’.

Subsequent research by Swisher and Hirsh [111] replicated the results of Efron [31]. However, this research also revealed that the temporal deficit in LHD patients is modulated by whether or not the stimuli are spatially separated. Furthermore, Swisher and Hirsh [111] found evidence for an involvement of the right hemisphere in temporal sequencing. A group of right hemisphere damaged (RHD) patients performed normally at visual presentations or when both tones of a pair were presented to the right ear. However, performance dropped when one of the tones of a pair was presented to the left ear.

Although drawing a link between time and speech, this early work made no use of speech stimuli. In order to establish such a link, it is useful to investigate patients’ ability to perceive and produce temporal information in speech. This has been done in a number of studies most of which investigated temporal production rather than perception. This imbalance between production and perception studies reflects the fact that it is relatively easy to measure the timing of speech production in patients and controls. In contrast, the investigation of temporal perception is relatively difficult as speech and prosody incorporate other acoustic parameters (e.g., frequency) that give additional and often redundant information as to the nature of the utterance. Therefore, although a time perception deficit might be present it can be obscured by the reliance on this additional information.

As indicated in the Introduction to this paper, speech is a very complex communication tool and therefore provides a rich ground for perceptual and acoustic analysis. For presentation purposes, the various measurements that have been performed will be categorized into four levels of analysis: a global level that investigates more general aspects of speech such as tempo\(^1\), which is often indicated by vowel, syllable or utterance duration; a syntactic level that investigates the translation of syntactic structure into prosodic phrasing and vice versa; a lexical-semantic level concerned with different forms of stress; and a segmental level that looks at consonant and vowel contrasts.

2.1. Global level

2.1.1. Perception

Hemispheric lateralization of the perception of speech tempo as a marker for the affective state of a speaker was first investigated by Van Lancker and Sidtis [118]. The authors presented sentences spoken with angry, sad, happy or surprised prosody. By means of discriminant function analysis they investigated which of the prosodic parameters—F0, amplitude or duration—could predict the identification errors of LHD and RHD patients. While LHD and

---

\(^1\) Research on emotional prosody is listed in the section dedicated to the global level of analysis because durational aspects of emotional prosody are commonly characterized by speech tempo, a global-level factor [102].
RHD patients showed a similar behavioral deficit in identifying the different prosodic expressions, the analysis performed on their errors indicated group differences in the reliance on acoustic cues. RHD patients seemed to base their judgment primarily on temporal cues, whereas LHD patients seemed to rely on F0. This finding led the authors to propose that the different prosodic parameters might be lateralized independently in the brain.

In an attempt to replicate the results from Van Lancker and Sidtis [118], Pell and Baum [92] performed a similar analysis on the errors of lesion patients in identifying different emotional (angry, sad, happy) and linguistic (declarative, interrogative, imperative) prosodic expressions. The authors found no evidence that LHD and RHD patients rely on different prosodic parameters and thus concluded that despite performance differences basic acoustic processing is comparable between LHD and RHD patients. Note, however, that only F0 differentiated significantly between the three emotional and the three linguistic speech categories. Moreover, discriminant analysis of the stimuli revealed only one significant function, which was related to fundamental frequency. This indicates that their speaker used mainly F0 to express emotional and linguistic meaning. Hence, it is not surprising that F0 was the only cue that could predict the patients’ errors.

2.1.2. Production

Ryalls [99] investigated the timing of speech production following cortical insult. He asked Broca’s aphasics and normal speakers to repeat a target sentence for which he measured vowel, word, and sentence duration as well as the F0 range. Acoustic analysis revealed that the slower speech in patients correlated with longer vowel durations. Furthermore, Ryalls [99] reported a significantly restricted F0 range for patients as compared to normal speakers. As Ryalls [99] investigated patients with left hemisphere lesions only, this research leaves open whether the reported deficits are specific to LHD or whether they can also be found in patients with RHD.

That brain damage to the right hemisphere can impair the production of prosody is suggested by the work of Blonder et al. [15]. They analyzed the spontaneous speech of a woman before and after she suffered a stroke affecting right fronto-temporo-parietal regions and the right basal ganglia. In contrast to the pre-stroke speech sample, the post-stroke sample was more restricted in F0 range; it was spoken faster and showed less variability in pause duration. Furthermore, Pell [90, 91] observed deficits in the production of emotional intonation in a sample of RHD patients. Compared to healthy speakers, their F0 and speech rate distinguished less between the different emotional categories (i.e., neutral, happy, angry, sad).

Together these findings suggest that more general aspects of speech prosody might be affected from damage to the left or the right cortical hemisphere and that both hemispheres might be important for aspects of speech prosody that are not tied to a specific linguistic function (e.g., intonation, stress). Note, however, that the studies cited in this context do not unambiguously implicate the left or right cortical hemisphere in prosodic production. It is possible that a disruption of other, closely related linguistic or emotional functions underlies the observed deficits. Moreover, with respect to the Blonder et al. [15] study, it is possible that subcortical damage accounted for the patients abnormal prosody.

2.2. Syntactic level

2.2.1. Production

Danly and Shapiro [26] looked at the prosodic realization of syntactic structure indicated by prosodic cues such as the terminal falling contour and the segmental lengthening of words in sentence final position. They found that a group of Broca’s aphasics modulated F0 to indicate sentence offset similar to healthy speakers. Sentence final lengthening, however, was absent in these patients. Gandour et al. [43] also reported abnormal timing at the sentence level following insult to the left hemisphere. Compared to RHD patients and healthy controls, LHD patients made longer pauses between words and marked the syntactic structure of a sentence less clearly via temporal cues.

Despite the reported difficulties in timing at the syntactic level, research by Schirmer et al. [105] suggests that certain aspects of prosodic phrasing are preserved following LHD. Specifically, LHD and RHD patients produced normal lengthening and pauses to indicate syntactic boundaries within a sentence. Further evidence for a preserved ability to prosodically convey syntactic structures following damage to the left hemisphere comes from Baum et al. [11]. Nevertheless, both Baum et al. [11] and Schirmer et al. [105] reported some deficits in the temporal sentence structure of their LHD patients. In congruence with other research [26, 43], they found longer pauses within sentences and the omission of sentence final lengthening.

A further syntactic/prosodic phenomenon that has been investigated in brain-damaged patients is syntactic focus. Schirmer et al. [105] reported an intact temporal implementation of focus in patients with left or right hemisphere lesions. They presented questions such as ‘Who promised to help Anna?’ and asked participants to respond by reading the corresponding sentence ‘Peter promised to help Anna and to clean the office’ with stress on the questioned constituent (i.e., Peter). While LHD and RHD patients lengthened the questioned constituent and inserted a subsequent pause, they failed to modulate F0 appropriately.

Ouellette and Baum [87] investigated a closely related phenomenon: contrastive stress. Speakers use contrastive stress to indicate the correct of two alternatives. For example, in normal speech the question ‘Who took the bus: the man or the woman?’ may be answered with a sentence like ‘The man took the bus’ with stress placed on ‘man’—the correct alternative. Ouellette and Baum [87] found that LHD
patients failed to use temporal cues to mark contrastive stress correctly, whereas measures of F0 and amplitude revealed no deficit. In contrast to LHD patients, RHD patients performed normally with respect to all three prosodic parameters. Based on these results, Ouellette and Baum [87] concluded that LHD may not induce a general prosodic impairment but rather a specific deficit in timing. They furthermore argue that, although the previously formulated assumption of independent lateralization of prosodic parameters [118] could explain their findings, it is possible that temporal information is programmed subcortically and that LHD patients simply fail to implement this information.

2.3. Lexical-semantic level

2.3.1. Perception

Baum [7] investigated the perception of temporal and F0 cues that indicate word stress. She presented utterances that differed only in stress placement such as ‘GREEN house’ and ‘green HOUSE’. These utterances either contained all the acoustic cues available in speech or were neutralized for duration or F0. Baum [7] found that patients with left hemisphere lesions were unable to decide between the two possible meanings with better-than-chance accuracy, even with the full set of acoustic cues. Patients with right hemisphere lesions performed significantly better than their left hemisphere counterparts when all the acoustic cues were available. However, their performance was just as impaired when one of the acoustic cues was neutralized. Unfortunately, there was only a small difference in syllable duration between the presented stress pairs so that even healthy controls performed at chance level for the F0 neutralized stimuli. Therefore, this study allows no conclusions about the lateralization of temporal processing. However, it suggests that LHD might disrupt the processing of lexical-semantic information conveyed by F0.

2.3.2. Production

A speech production study by Ouellette and Baum [87] revealed that temporal cues of word stress are compromised following LHD but not RHD. Moreover, they reported that the stress production deficit found in LHD patients was restricted to timing as other cues such as F0 and amplitude were preserved. Further evidence for LHD resulting in impaired stress production comes from Gandour and Baum [38]. Their study investigated stress clash, which reflects the preference to maintain an alteration of stressed and un-stressed syllables in spoken English. Because of this preference, stress is moved in words like ‘thirTEEN’ to word initial position when immediately followed by words that also carry word initial stress (e.g., ‘THIRteen WOmen’). In accordance with previous work [87,90,91], Gandour and Baum [38] reported normal performance in RHD patients but stress clash impairments in patients with LHD. Interestingly, these impairments showed not only for timing but also for the F0 contour, which was marked by continuation rises. As these continuation rises seemed to be due to increased pause durations, Gandour and Baum [38] proposed that the F0 deficit in LHD patients was only secondary to a primary temporal deficit.

2.4. Segmental level

2.4.1. Perception

One of the best studied temporal phenomena at the segmental level is the VOT of stop consonants. A typical task used to investigate VOT comprises words or syllables that start with a stop consonant (e.g., ‘bath’ vs. ‘path’). Speech synthesis is often used to create stimuli that vary on a VOT continuum between the two endpoints: voiced and voiceless. If normal listeners are asked to label speech synthetic stimuli, they do not show a continuous increase in ‘voiceless’ responses as the VOT increases. Rather, there is a short transition from ‘voiced’ to ‘voiceless’ responses indicating a category boundary. This categorical perception of VOT seems to be impaired both in fluent and non-fluent aphasics [8,16,39,49]. Moreover, in some cases these patients are unable to identify even the endpoint stimuli of the VOT continuum. If the sentence or other contextual information biases one of the two opposing stop categories, LHD patients base their decision to a greater extent on this additional information than do normal listeners [8,49]. Interestingly, performance of RHD patients in VOT tasks is comparable to that of healthy listeners suggesting that the left but not the right hemisphere is critical for VOT discrimination [16,39].

2.4.2. Production

Similar to perception, the production of VOT is compromised in aphasia. Early evidence comes from Gandour and Dardarananda [40] who asked a sample of Thai speakers to produce words with an initial consonant of bilabial (e.g., ‘baan’), alveolar (e.g., ‘dam’) or velar (e.g., ‘kan’) place of articulation and that had either a short or long VOT (e.g., ‘baan’ vs. ‘paan’). The measurement of VOT in nonfluent global and Broca aphasics revealed a dramatic overlap of the VOT distributions of opposing stop categories, which means that these patients failed to differentiate temporally between consonants that had the same place of articulation. In contrast, VOT production of patients with fluent aphasia (i.e., conductive and Wernicke aphasia) was only mildly impaired. Furthermore, one patient with RHD produced VOTs that were indistinguishable from the VOTs of healthy speakers. These findings are congruent with many other studies of VOT production following cortical insult [9,10,17,42].

\[2\] Note, that Blumstein et al. [18] found an increased influence of heuristic strategies on VOT perception only in patients with nonfluent but not with fluent aphasia. Baum [8], however, demonstrated an increased influence of sentence context on VOT perception in both nonfluent and fluent aphasics.
Ryalls et al. [102], however, reported only mild deficits in the production of VOT in French-speaking Broca’s aphasics. They argue that the greater average VOT difference between voiced and unvoiced pairs found in French might explain these controversial findings. Furthermore, Kurowski et al. [68], found evidence that although VOT of stop consonants might be preserved in RHD patients, the production of other types of segmental information can be impaired. The authors reported that RHD patients had difficulties differentiating between voiced and unvoiced fricatives (e.g., ‘v’ vs. ‘f’) on a temporal scale. In normal speech, unvoiced fricatives are about 40 ms longer than voiced counterparts [65]. Nevertheless, the segmental deficit in RHD patients was less severe than the deficit that has been reported for LHD patients [10].

In addition to the temporal characteristics of consonant production, some investigators looked at the timing of vowels in brain-damaged populations. While it has been reported that aphasics produce generally longer vowels than healthy speakers [10,101] it seems that relative timing is preserved. Gandour and Dardarananda [41], for example, investigated vowel length in Thai where it is used to distinguish between different meanings of otherwise comparable words. For example, the word ‘ha’t spoken with a short vowel means ‘to practice’, whereas the word ‘haat’ spoken with a long vowel means ‘shoal’. To investigate whether this contrastive function of vowel length is disrupted in aphasia, Gandour and Dardarananda [41] asked their participants to read words containing short and long vowels and measured vowel duration. Acoustic analysis revealed a comparable performance of aphasic patients, a patient with a right hemisphere lesion, and healthy controls. That the contrastive function of vowel length is spared following LHD has been reported by other researchers as well [10] and has been explained by the lower level of articulatory control required for vowel as compared to consonant production. Alternatively, it has been attributed to the different temporal processing demands of vowels as compared to consonants [41]. It has been hypothesized that the left hemisphere is specialized for the processing of durations up to 50 ms, whereas longer durations, such as for vowels, would be processed by either hemisphere [83].

### 2.5. Summary

An overview of the studies that have been cited in the preceding text is presented in Tables 1 and 2. They are categorized according to the level of speech at which temporal processing has been investigated. Note, however, that this categorization does not imply that these levels are independent of each other. During speech production and perception, temporal information has to be integrated across the different levels. Therefore, it is possible that difficulties at one level have a negative effect on the temporal processing at another level. Note also that the tables do not represent an exhaustive review of the literature concerned with the timing of speech perception and production. The present paper merely aims at providing an overview of important findings and outlining the commonalities in order to allow some conclusions about the underlying neuroanatomical structures that support temporal processes.

One conclusion that can be drawn from this overview is that lesions in the right hemisphere are less disruptive to the timing of speech than are lesions in the left cortical hemisphere. One might argue that patients with RHD have not been studied as extensively as patients with LHD.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Perception studies in patients with cortical lesion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of analysis</td>
<td>Study</td>
</tr>
<tr>
<td>Global level</td>
<td>[92]</td>
</tr>
<tr>
<td>Segmental level: consonants</td>
<td>[8]</td>
</tr>
<tr>
<td></td>
<td>[18]</td>
</tr>
<tr>
<td></td>
<td>[16]</td>
</tr>
<tr>
<td></td>
<td>[39]</td>
</tr>
<tr>
<td></td>
<td>[49]</td>
</tr>
<tr>
<td>= normal performance; – = deficient performance (in case of several groups being impaired their deficit is rank ordered); no entry = this group has not been studied; F = fluent aphasia, NF = non-fluent aphasia, B = Broca aphasia, TM = transcortical motor aphasia, C = conduction aphasia, G = global aphasia, W = Wernicke aphasia.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Production studies in patients with cortical lesions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of analysis</td>
<td>Study</td>
</tr>
<tr>
<td>Global level</td>
<td>[15]</td>
</tr>
<tr>
<td></td>
<td>[92]</td>
</tr>
<tr>
<td></td>
<td>[101]</td>
</tr>
<tr>
<td></td>
<td>[26]</td>
</tr>
<tr>
<td></td>
<td>[42]</td>
</tr>
<tr>
<td></td>
<td>[107]</td>
</tr>
<tr>
<td>Lexical-semantic level</td>
<td>[38]</td>
</tr>
<tr>
<td></td>
<td>[88]</td>
</tr>
<tr>
<td></td>
<td>[92]</td>
</tr>
<tr>
<td></td>
<td>[10]</td>
</tr>
<tr>
<td></td>
<td>[9]</td>
</tr>
<tr>
<td></td>
<td>[16]</td>
</tr>
<tr>
<td></td>
<td>[17]</td>
</tr>
<tr>
<td></td>
<td>[40]</td>
</tr>
<tr>
<td></td>
<td>[43]</td>
</tr>
<tr>
<td></td>
<td>[68]</td>
</tr>
<tr>
<td></td>
<td>[102]</td>
</tr>
<tr>
<td>Segmental level: consonants</td>
<td>[10]</td>
</tr>
<tr>
<td></td>
<td>[41]</td>
</tr>
<tr>
<td></td>
<td>[68]</td>
</tr>
<tr>
<td>+ = normal performance; – = deficient performance (in case of several groups being impaired their deficits are rank ordered); no entry = this group has not been studied; F = fluent aphasia, NF = non-fluent aphasia, B = Broca aphasia, TM = transcortical motor aphasia, C = conduction aphasia, G = global aphasia, W = Wernicke aphasia.</td>
<td></td>
</tr>
</tbody>
</table>
However, out of the 18 cited studies (4 perception, 14 production) that included patients with RHD only 4 report impaired timing. Two of these studies investigated the production of VOT and reported a relatively mild deficit for RHD patients [42,68]. The other two studies found alterations in speech tempo [15,92], which is influenced not only by timing but also by other factors such as affect, arousal, speech motor control, and verbal fluency.

Compared to RHD patients, LHD patients seem severely impaired during both the perception and production of duration at various levels of speech. Not only do they show alterations in the perception [120] and production of speech tempo [101], they also have difficulties in indicating sentence offset via sentence final lengthening (e.g., Ref. [107]). They fail to produce clear durational patterns that help to distinguish contrastive and phonetic stress pairs [88] and they exhibit deficient perception and production of VOT (e.g., Refs. [39,40]). Most studies that investigated this latter point also looked at performance differences between different types of aphasia. It appears that non-fluent aphasics with lesions in the anterior part of the left hemisphere are more severely impaired in the production of VOT than fluent aphasics with posterior lesions. More specifically, anterior patients show a large overlap in the VOT of voiced and voiceless stops. This deficit has been classified as phonetic and contrasted to the phonemic errors made by posterior patients, which show no overlapping VOT distributions, but produce the phonemic errors made by posterior patients, which lead to an over- or underestimation of time intervals, respectively [79,80]. Based on this, it has been proposed that the BG play a central role in timing. Moreover, it has been suggested that the internal clock or pacemaker is dependent upon the integrity of cortico-striatal circuits (e.g., Refs. [47,80]).

Given these findings it seems that both anterior and posterior damage to the left hemisphere can disrupt temporal aspects of speech. What remains unclear, however, is whether the observed temporal deficits do indeed reflect impaired temporal processing. It is possible that they arise as a secondary effect from impairments in the control of the articulatory effectors or deficient representations of linguistic information. That this is an important concern becomes clear from studies that report additional deficits in the processing of other prosodic parameters such as the F0 contour (e.g., Ref. [107]). Therefore, this concern should be kept in mind when drawing any conclusions about the left hemisphere’s role in timing speech.

3. Patients with basal ganglia dysfunction

The basal ganglia (BG) consist of four subcortical nuclei: the striatum (i.e., caudate nucleus, putamen), the globus pallidus, the substantia nigra, and the subthalmic nucleus. Furthermore, the BG receive input from various regions of the cortex suggesting an involvement in a variety of cognitive and motor functions. Several neurological disorders such as Parkinson’s and Huntington’s disease (PD, HD, respectively) have been related to BG dysfunction. Besides deficits in motor control, these diseases are characterized by alterations in speech production [27,28]. For example, the speech of PD patients has been described as flat and monotonous, imprecise with respect to consonant production, and somewhat hastened. Additionally, there is evidence for an impaired perception of speech prosody. PD patients have difficulties recognizing the emotional and the linguistic information that is expressed prosodically. Based on the finding of normal performance in simple discrimination tasks (e.g., same–different judgment of two successive utterances) it has been argued that BG dysfunction impairs the mapping of prosodic features onto corresponding linguistic and affective representations leaving low-level prosodic perception undisturbed [73,90]. Note that this research did not specify which prosodic features PD patients fail to map onto corresponding linguistic and affective representations.

Another function that has been found compromised in BG dysfunction is interval timing. In an unmedicated state, PD patients reproduce a previously learned duration less accurately than age-matched controls [75]. Furthermore, it has been demonstrated that the administration of dopaminergic drugs, which modulate BG function, can increase or slow down the speed of an internal clock or pacemaker, which leads to an over- or underestimation of time intervals, respectively [79,80]. Based on this, it has been proposed that the BG play a central role in timing. Moreover, it has been suggested that the internal clock or pacemaker is dependent upon the integrity of cortico-striatal circuits (e.g., Refs. [47,80]).

The following section reports studies that investigated whether BG dysfunction compromises the processing of time in speech. It seems reasonable to assume that the temporal deficits found in interval timing studies extend to the temporal processing of speech. For example, changes in clock speed reported in PD patients [89] might affect aspects of speech that require absolute timing such as the perception of speech tempo. This could compromise the recognition of emotions that are marked by slow and fast speech. For example, PD patients may not perceive a slow speaking pace as slow and thus may not easily recognize the sad emotion that is often associated with slow speech.

Note that except for a few studies that failed to provide medication information [34,121], the studies listed below were conducted while patients were on medication and hence may not reveal the full range of temporal processing deficits exhibited by unmedicated PD patients (e.g., Ref. [75]). If possible, the following text provides information about the stage of the disease according to Hoehn and Yahr [124].

3.1. Global level

3.1.1. Perception

To investigate whether BG dysfunction compromises the perception of temporal cues in speech, Breitenstein et al. [20] asked a group of PD patients (HY: 19 stage II and 1
stage (IV) to identify sentences with happy, sad, angry, frightened, and neutral prosody. They manipulated either the F0 variation or the speaking rate of these sentences so that both parameters increased linearly from 50% to 150% of their original value. Identification scores revealed a similar influence of the pitch manipulation on emotional perception in healthy controls and PD patients. Similar results were obtained for the temporal manipulation with a few exceptions. The speech tempo of sad utterances needed to be slower for PD patients than for normal listeners in order to be identified correctly. Furthermore, the PD group misclassified slow stimuli more frequently as frightened than did the control group. One possible explanation for these effects is that the internal clock of PD patients was slowed down leading to an underestimation of time. Durations might have been perceived as shorter than they were, which would explain the failure of PD patients to recognize slow speech as slow.

3.1.2. Production

Stereotaxic thalamotomy or electrical stimulation of the ventrolateral thalamic nucleus has been found to abolish tremor in PD. Interestingly, these treatments do not only affect tremor but also temporal aspects of speech production [103]. Canter and Van Lancker [24], for example, report a case of speech acceleration following bilateral thalamic surgery. Additionally, the authors noticed a distortion of the temporal relationship between vowels and consonants which affected the patient’s speech intelligibility. This suggests that the ventrolateral nucleus of the thalamus and connections between the thalamus and the basal ganglia contribute to temporal aspects of speech production.

The importance of neuronal circuits connecting thalamus and basal ganglia is indicated by findings of impaired temporal speech patterns in PD patients with non-lesioned thalamic nuclei. This group of patients has also been reported to show speech acceleration [30], which in some thalamic nuclei. This group of patients has also been reported to show temporal speech patterns in PD patients with non-lesioned and basal ganglia is indicated by findings of impaired connections between the thalamus and the basal ganglia suggest that the ventrolateral nucleus of the thalamus and which affected the patient’s speech intelligibility. This suggests that the ventrolateral nucleus of the thalamus and connections between the thalamus and the basal ganglia contribute to temporal aspects of speech production.

The importance of neuronal circuits connecting thalamus and basal ganglia is indicated by findings of impaired temporal speech patterns in PD patients with non-lesioned thalamic nuclei. This group of patients has also been reported to show speech acceleration [30], which in some cases might approach a rate of 250 words per minute [23]. Note, however, that not all PD patients show accelerated speech and that some might even tend to speak slower than normal ([50]; HY: 2 stage I, 7 stage II, 1 stage III). They presented words that varied on a CLT continuum (i.e., ‘Boden’ vs. ‘Boten’; translation: floor vs. messengers) or on a VOT continuum (i.e., ‘tick’ vs. ‘dick’; translation: tick vs. fat). No differences between patients and healthy controls were obtained in the VOT perception task. In the CLT perception task, 3 out of 10 PD patients always perceived the word ‘Boden’ and failed to recognize longer CLTs as ‘Boten’. In a second experiment, the authors extended the CLT continuum and presented words with CLTs that exceeded 180 ms. The finding that the three PD patients could perceive the word ‘Boten’ for the newly added longer CLT trials was interpreted as reflecting an underestimation of time in these patients. That all PD patients performed normally in the VOT task was attributed to them relying on additional information that distinguishes between word initial stop consonants (i.e., loudness of the aspiration phase). This additional information was not available to help identify voiced and unvoiced stops in word medial position.

3.3. Segmental level

3.3.1. Perception

Gräber et al. [50] investigated the perception of segmental information in patients with PD (HY: 2 stage I, 7 stage II, 1 stage III). They presented words that varied on a CLT continuum (i.e., ‘Boden’ vs. ‘Boten’; translation: floor vs. messengers) or on a VOT continuum (i.e., ‘tick’ vs. ‘dick’; translation: tick vs. fat). No differences between patients and healthy controls were obtained in the VOT perception task. In the CLT perception task, 3 out of 10 PD patients always perceived the word ‘Boden’ and failed to recognize longer CLTs as ‘Boten’. In a second experiment, the authors extended the CLT continuum and presented words with CLTs that exceeded 180 ms. The finding that the three PD patients could perceive the word ‘Boten’ for the newly added longer CLT trials was interpreted as reflecting an underestimation of time in these patients. That all PD patients performed normally in the VOT task was attributed to them relying on additional information that distinguishes between word initial stop consonants (i.e., loudness of the aspiration phase). This additional information was not available to help identify voiced and unvoiced stops in word medial position.

3.3.2. Production

Research on speech production suggests that patients with mild PD (HY: stage I and II) use temporal information to distinguish between speech segments. Ackermann et al. [2], for example, found no differences between patients with mild PD (HY: 5 stage I, 4 stage II) and healthy controls in the production of vowel length and VOT contrasts. Lieberman et al. [70], found a deficit in only 2 out of 20 patients with mild PD (HY: stage I and II) and in 7 out of 20 patients...

---

3 The authors failed to indicate the stage of the disease of these patients.
with moderate PD (HY: stage III). Similar to what has been reported for Broca’s aphasics, the VOT distributions for voiced and unvoiced stop consonants overlapped in these patients. Because of the similarities in speech pathology between patients with Broca’s area lesioned and patients with BG dysfunction, the authors argued that the neuronal basis for these speech impairments is a disconnection of Broca’s area from other regions in the prefrontal cortex. Moreover, they proposed that PD affects the cortico-striatal circuits [6] that are supposed to mediate the communication between different speech relevant regions in the frontal cortex.

3.4. Summary

Table 3 provides an overview of the studies that investigated speech timing in BG dysfunction. Deficits in the perception of temporal information have been reported for global aspects of speech such as emotional prosody [20]. More specifically, PD patients seem to underestimate time, which leads to difficulties in the perception of emotions such as sadness or fright that are marked by slow and fast speech, respectively. Similar perceptual deficits have been reported at the segmental level. Some PD patients require longer than normal CLTs to perceive unvoiced word medial stops [50]. Together these data from speech perception suggest an underestimation of time or a slowing of the internal clock in PD.

In addition to the perception of temporal information in speech, the BG seem to contribute to the timing of speech production. For example, there are reports of abnormally fast speech in BG dysfunction (e.g., Ref. [23]). Furthermore, BG dysfunction reduces the number and length of speech pauses [34] and interferes with the ability to accelerate speech production voluntarily [121]. In contrast to the speech perception deficits, these findings cannot easily be explained with changes in clock speed, which suggests that other aspects of timing might be disturbed as well.

4. Patients with cerebellar lesions

Early work by Ivry and Keele [59] renewed interest in the cerebellum as a candidate structure that is critical for timing. The authors reported that patients with cerebellar lesions performed worse than healthy controls when asked to discriminate between time intervals. Based on these and other findings, Keele and Ivry [63] proposed that the cerebellum serves as an internal clock for motor and perceptual temporal computations.

4.1. Global level

4.1.1. Production

Clinical observations of patients with cerebellar dysfunction suggest that the cerebellum is involved in temporal aspects of speech. Patients who suffer from cerebellar dysfunction often show a speech motor disorder termed ataxic dysarthria [27,28]. Darley et al. [27,28] describe the speech in patients with ataxic dysarthria as having errors in timing, force, range, and direction of speech movements. Similar to hypotonic dysarthria found in PD, ataxic patients show imprecise consonant production. Additionally, speech production is slowed and speech rhythm is equalized [27,55]. The effect of equalized speech rhythm is also referred to as ‘scanning speech’ and is due to patients with ataxic dysarthria tending to produce syllables of equal duration within an utterance or breath group. This tendency reflects deficiencies in the shortening of unstressed syllables and is thought be an articulatory rather than a prosodic problem [55]. In addition to the decrease in duration variability within an utterance, there are other acoustic measures that show an increase in variability such as the syllable duration across utterances and the duration of inter-stress intervals. As healthy speakers tend to keep an equal distance between the onset of successive stress groups, this latter effect additionally contributes to the perceived lack of rhythm in ataxic dysarthria [55].
4.2. Segmental level

4.2.1. Perception

To determine whether temporal deficits can be found in speech perception, Ackermann et al. [1] conducted a study with cerebellar patients that is similar to their investigation of speech perception in PD [50]. They presented words that varied on a closure time continuum (i.e., ‘Boden’ vs. ‘Boten’) or on a voice-onset time continuum (i.e., ‘tick’ vs. ‘dick’). Patients with lesions in only one cerebellar hemisphere performed similarly to normal participants, whereas patients with both cerebellar hemispheres lesioned failed to produce a significant phoneme-boundary effect for the CLT stimuli. In contrast, the perception of VOT stimuli was preserved, which was attributed to the presence of other acoustic cues that differentiate between voiced and unvoiced word initial stops.

In a further study, Ackermann et al. [5] investigated the perception of non-speech stimuli (i.e., two tone sequences) that were comparable to speech in their durational characteristics. In one task, they asked patients with cerebellar atrophy to indicate whether the first or the second duration of a pair was longer. In a second task, patients had to classify pauses as either short or long. They found that cerebellar patients performed worse than normals in the first task, whereas performance in the second task was comparable between groups. Based on their findings, the authors argued that temporal perception of speech-like stimuli is preserved in cerebellar atrophy and that the deficit obtained in the first task for comparing durations resulted from impaired working memory rather than from disturbed temporal processing. Moreover, they explained their previous finding of impaired CLT perception in cerebellar patients [1] as due to higher computational demands for time intervals that are embedded in words.

4.2.2. Production

Similar deficits have been reported during speech production. Ackermann et al. [2,3] found that cerebellar patients have difficulties in the contrastive use of VOT and vowel length. However, whether or not patients show these production deficits seems to depend on speech motor coordination demands. Disrupted timing is present only when several articulatory organs (e.g., tongue, lips, velum) need to be coordinated.

4.3. Summary

Patients with cerebellar lesions frequently suffer from a speech disorder termed ataxic dysarthria. One prominent feature of this disorder is the disruption of temporal control as evidenced by an equalized speech rhythm [27]. Attempts to corroborate these clinical observations with experimental investigations of temporal speech processes in cerebellar patients are rare (see Table 4). Although the available evidence implicates the cerebellum in temporal aspects of speech, it seems that cognitive and motor coordination demands modulate the involvement of the cerebellum in speech timing tasks. Thus, it has been suggested that an interaction between the cerebellum and speech relevant cortical regions underlies the perception and production of temporal information in speech. Moreover, these proposed cortico-cerebellar circuits are thought to be recruited only under high cognitive or motor coordination demands. The comparison of durations stored in working memory or the need to control several articulatory organs represents such demands [2,5].

5. Studying the normal brain

Although patient studies can be indicative of which brain regions contribute to certain cognitive functions, they face several limitations. For example, patient groups are often heterogeneous with respect to the type of disease (tumor, occlusive vs. hemorrhagic strokes, atrophy, etc.) and the exact location of the neuronal insult. A further problem concerns the interpretation of deficits that patients might show in a specific task. In addition to the process under study (e.g., timing), experimental tasks always involve other aspects of cognition and behavior that might be impaired due to brain damage and affect task performance. Brain imaging methods such as event-related potentials (ERPs), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and positron emission tomography (PET) are therefore helpful in verifying and adding to the knowledge obtained in patient studies. The following sections present research that made use of these techniques in order to determine the neuroanatomical regions that support the perception of temporal information in speech.

| Table 4 |
| Production and perception studies in patients with cerebellar lesions |
| Level of analysis | Modus | Study | Results |
| Global level | Production | [55] | – longer syllable durations and stress intervals, reduced variability of intra-utterance syllable durations, increased variability of inter-utterance syllable durations and of inter-stress interval durations |
| Segmental level | Perception | [1] | – patients with both cerebellar hemispheres lesioned failed to show phoneme boundary effects for CLT stimuli; normal phoneme boundary effects for VOT stimuli |
| | | [5] | – normal performance when classifying pauses as long or short; deficits when indicating whether the first or the second stimulus of a pair was longer |
| | Production | [3] | – deficits in the production of voiced and unvoiced stop consonants |
| | | [2] | – deficits in the production of VOT and vowel contrasts |
5.1. Global level

There is little research investigating the regions involved in the processing of global temporal aspects of speech such as tempo. Pihan et al. [97] made a start by comparing the perception of rate- and pitch-manipulated sentences in order to disentangle areas involved in the processing of durational and F0 information. Their sentences had happy, neutral, or sad intonation and varied in expressiveness depending on whether the speech tempo was fast, medium, or slow and the F0 range was large, medium, or small. Participants were asked to indicate the emotion conveyed by each sentence pair and to decide which of the two sentences was more expressive. In a first experiment participants performed the task without using inner speech and in a second experiment participants were instructed to repeat sentences subvocally. The results failed to reveal differences between rate- and pitch-manipulated sentences in the topographical distributions of the neuronal activity as measured with the EEG. For both types of stimuli, neuronal activity was lateralized to left frontal regions when subjects repeated sentences using inner speech, whereas activity was lateralized to right frontal regions in the absence of inner speech.

5.2. Rhythm

There seems to be no study that investigates the neuroanatomical structures that support the processing of durational parameters of stress and phrasing in normal listeners. However, research has been conducted on the perception and production of rhythm—a musical phenomenon that is closely related to the metrical properties of speech.

Penhune et al. [95] presented tone sequences and sequences of squares with a certain rhythm and asked participants to repeat that rhythm by pressing and releasing a response key. They found that regardless of input modality the cerebellum was more strongly activated when the rhythm altered in each trial than when it stayed the same. Because an alteration in rhythm required the participants to alter the timing of their response, the authors argued that their findings indicate the importance of the cerebellum for motor timing. The failure to find an increase in BG activity for the alternating rhythms was attributed to the BG being equally involved in the production of alternating and repeating rhythms.

Riecker et al. [100] performed a similar study in which they asked participants to produce the syllable sequence ‘papapa’ with either an isochronous or a rhythmical pattern (e.g., short—short—long). Compared to the isochronous condition, the rhythmical condition activated putamen and thalamus in the left hemisphere and superior temporal gyrus, temporal plane and premotor cortex in the right hemisphere. In contrast to Penhune et al. [95], the authors found activation in the cerebellum only when speech production was compared with a passive listening baseline.

Schubotz et al. [108] reported activity in both the cerebellum and the BG for the perception of visual and auditory rhythms. Additionally, they found bilateral activity in the premotor cortex, the frontal operculum, the intraparietal sulcus, and in the supplementary motor area. They obtained these activations by comparing the detection of visual and auditory rhythmical deviants with the detection of color and tone deviants. Thus it seems that different cortical regions in both hemispheres, the BG and the cerebellum contribute to the perception and production of rhythms.

5.3. Segmental level

Before reviewing the literature on consonant and vowel duration processing, studies that examined the temporal processing of tones shall be presented. As tones share some commonalities with speech, comparing research on temporal processing of tones and speech might provide useful insights.

Preattentive processing of tones has been investigated with ERPs by means of the mismatch negativity paradigm. In this paradigm, participants usually hear a series of standard tones that is occasionally interrupted by a deviant that differs from the standard in a single aspect such as frequency. Although participants are engaged in a different task (e.g., reading), the ERP indicates that they nevertheless process standards and deviants differently. Deviants elicit a larger negativity (mismatch negativity; MMN) between 100 and 250 ms following stimulus onset (see Ref. [85]). Furthermore, the ERP difference between the standard and the deviant is maximal over frontal electrodes and has been traced back to a bilateral source in the auditory cortex [66]. Similar to the ERP of frequency deviants, the ERP of duration deviants is characterized by an MMN [32]. Moreover, this mismatch effect is not restricted to the perception of tones of different length but can also be found for speech stimuli. In an MEG study, Ackermann et al. [4] found the magnetic analog of the MMN in response to CV deviants that differed from the CV standard in VOT. Although both hemispheres equally contributed to this mismatch effect it appeared earlier in the left than in the right auditory cortex suggesting a left hemisphere specialization in the processing of linguistically relevant durations.

In contrast to the preattentive bilateral mechanisms, attentive discrimination of tones that differ in duration seems to be lateralized to the right hemisphere. Belin et al. [13] conducted a PET study in which they found activity in the right inferior frontal operculum, the right inferior parietal lobe and the left cerebellum that was related to the discrimination of tone durations. Because similar regions were found to contribute to the discrimination of stimulus intensity [12], they were thought to be components of a network that directs attention to different types of sensory information. In contrast, the right cerebellum, the left thalamus, and the left and right striatum were specifically activated when listeners discriminated tone durations. Therefore, the authors propose that these latter regions represent components of a network concerned with the processing of temporal information only.
The bilateral cortical contributions to preattentive processing of tone durations and the right hemisphere dominance for attentive processing of tone durations seem to contradict evidence from patients with cortical lesions that suggests left hemisphere dominance. However, this evidence comes from studies that investigated the temporal processing of speech rather than non-speech stimuli. Therefore, the apparent discrepancy between neuroimaging and patient literature may simply reflect differences in the temporal processing of speech and non-speech stimuli. Two major hypotheses have been formulated with regard to such differences.

First, it has been argued that lateralization of temporal processing depends on the linguistic significance of a stimulus. In contrast to tones, speech stimuli have linguistic significance and therefore speech, but not tones, are processed mainly in the left hemisphere. Evidence for this assumption comes from an fMRI study by Gandour et al. [44]. The authors investigated the processing of vowel duration in Thai—a language in which changes in vowel duration can imply a change in meaning. Gandour et al. [44] presented Thai word pairs and pairs of non-speech stimuli (i.e., hums) that could differ in duration and asked participants to perform a same–different judgment. Compared to the non-speech condition, the speech condition elicited stronger activity in the left medial frontal gyrus suggesting that this region contributes to the processing of linguistically relevant durations. Interestingly, this left frontal activation was absent in Chinese listeners who did not speak Thai and therefore could not attribute any linguistic significance to vowel duration.

Further evidence for the assumption that lateralization of temporal processing depends on linguistic function comes from a study comparing duration discrimination in Thai words and hums in Thai and English listeners [45]. Thai but not English listeners showed activation in the left inferior frontal sulcus when discriminating vowel durations as compared to hum durations.

The second hypothesis that addresses hemispheric differences in the processing of speech and tones is based on acoustic differences between the two. In contrast to tones or music, speech incorporates more rapidly changing sound sequences. Especially the processing of stop consonants requires a temporal resolution within the range of tens of milliseconds. Therefore, it has been proposed that the left hemisphere’s specialization for speech reflects hemispheric differences in the time window over which acoustic information is integrated [83,109,114,124]. More specifically, the left hemisphere is thought to have a smaller temporal integration window and thus qualifies for the processing of rapidly changing acoustic events as present in speech.

First evidence for this assumption came from Mills and Rollman [83]. The authors conducted a dichotic listening study in which participants heard a sequence of tones that were separated by either 40, 50, 67, or 100 ms. The listeners had to indicate when the sequence was finished which required them to time the interval that separated the tones in a sequence. There was a right ear advantage (REA)/LH dominance for the 40- and 50-ms trains, whereas performance was comparable between ears for the longer durations. Based on their findings, Mills and Rollman [83] proposed that the left hemisphere might be specialized for temporal processing in the range of phonemic durations.

Similar conclusions were reached by Schwartz and Tallal [109]. The authors synthetically manipulated the rate at which formant information changed within CV syllables. They found that the REA/LH dominance, commonly reported for the dichotic processing of CV syllables [64], was influenced by their manipulation: the REA was smaller for extended formant transitions as compared to normal speech. More specifically, participants made more correct responses to left ear presentations and more incorrect responses to right ear presentation when formant transitions were synthetically extended so that the rate of acoustic change was slowed down as compared to normal speech.

Further probing into the hypothesis that rate of acoustic change underlies hemispheric specialization using PET revealed additional evidence. Compared to the detection of targets without rapidly changing information (i.e., steady-state vowels), the detection of targets with rapidly changing information (i.e., tone triplets, CV, and CVC syllables) activated the left frontal operculum more strongly [33]. Interestingly, stronger left frontal activity—albeit with somewhat different coordinates (i.e., left orbito-frontal cortex)—has also been reported for the discrimination between tones that contained fast as compared to slow changes in frequency [61]. This suggests that also the processing of

<table>
<thead>
<tr>
<th>Study</th>
<th>Tool</th>
<th>Comparison</th>
<th>Activated areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>[95]</td>
<td>PET</td>
<td>reproduction of alternating vs. repeating rhythms</td>
<td>BH: CB (visual and auditory input); anterior cingulate (visual input); SMA, putamen, red nucleus (auditory input)</td>
</tr>
<tr>
<td>[100]</td>
<td>fMRI</td>
<td>production of rhythmic vs. isochronous syllable trains</td>
<td>LH: putamen, thalamus</td>
</tr>
<tr>
<td>[108]</td>
<td>fMRI</td>
<td>detection of auditory/visual discordant vs. detection of tone/color deviants</td>
<td>BH: SMA, premotor cortex, frontal opercular cortex, intraparietal sulcus, putamen, cerebellum; RH: caudate nucleus</td>
</tr>
</tbody>
</table>

fMRI = functional magnetic resonance imaging; PET = positron emission tomography; CB = cerebellum; STG = superior temporal gyrus; TP = temporal plane; SMA = supplementary motor area; BH = both hemispheres.
tones can be lateralized to the left hemisphere, provided they contain rapidly changing information similar to speech. Further evidence for hemispheric differences in temporal coding comes from intracerebral recordings [71]. CV-syllables starting with voiced and unvoiced consonants elicit different evoked potentials in the left but not in the right Heschl’s gyrus and temporal plane. Moreover, the evoked potentials in the left auditory cortex seem to reflect the temporal structure of voiced and unvoiced CV syllables, which suggests a predestination of the left auditory cortex for the temporal coding of speech segments.

Taken together this research provides evidence for an influence of both linguistic and temporal stimulus properties on cortical lateralization. The processing of linguistically relevant stimuli and/or stimuli that incorporate fast acoustic changes seems to rely more strongly on left- than right-hemisphere mechanisms. More specifically, the left superior temporal cortex and the left frontal cortex have been shown to support the temporal processing of speech segments [21,78]. However, as the coordinates especially for the left frontal activation markedly differ between studies, exact neuroanatomical specification of the structures that support the timing of speech segments remains premature.

5.4. Summary

Tables 5 and 6 summarize the studies that investigated temporal processing in non-neurological participants. As is true for patients with neurological insults, the segmental level has been investigated most intensively. Respective research revealed a specialization of the left frontal and superior temporal cortex for temporal processing. However, whether or not these left hemisphere structures are more strongly engaged in processing than their right hemisphere homologues critically depends on the stimuli’s linguistic significance and/or their rate of acoustic change.

In contrast to the segmental level, little research has been conducted investigating other aspect of speech such as stress and speech tempo. However, the production of rhythm—a phenomenon closely related to stress—seems to recruit bilateral or right superior temporal and frontal cortex. Given the findings from the segmental level, one would assume

<table>
<thead>
<tr>
<th>Study</th>
<th>Tool</th>
<th>Comparison</th>
<th>Activated areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>[4]</td>
<td>MEG</td>
<td>passive listening to deviant CVs that differ from the standard in VOT</td>
<td>RH: mismatch negativity in the supratemporal plane (earlier in the left than in the right hemisphere)</td>
</tr>
<tr>
<td>[13]</td>
<td>PET</td>
<td>detecting tone deviants that are longer than the standard vs. passive listening to the standard</td>
<td>LH: CB</td>
</tr>
<tr>
<td>[21]</td>
<td>fMRI</td>
<td>same—different judgment of word pairs that could differ in VOT vs.</td>
<td>RH: inferior frontal operculum, orbital prefrontal cortex, putamen, caudate nucleus, IFG, MFG, MTG, inferior parietal cortex, CB</td>
</tr>
<tr>
<td>[33]</td>
<td>PET</td>
<td>auditory target detection (tone triplets, vowels, CV and CVC syllables) vs. visual fixation</td>
<td>BH: STG</td>
</tr>
<tr>
<td>[32]</td>
<td>EEG</td>
<td>passive listening to tone deviants that are longer than the standard</td>
<td>LH: MFG/SMA, thalamus, CB</td>
</tr>
<tr>
<td>[44]</td>
<td>fMRI</td>
<td>same—different judgment of Thai monosyllabic word pairs that could differ in duration vs. same—different judgment of ham pairs that could differ in duration same—different judgment vs. passive listening to ham pairs that could differ in duration</td>
<td>LH: frontal operculum</td>
</tr>
<tr>
<td>[45]</td>
<td>fMRI</td>
<td>same—different judgment of Thai monosyllabic word pairs that could differ in duration vs. same—different judgment of ham pairs that could differ in duration same—different judgment vs. passive listening to ham pairs that could differ in duration</td>
<td>BH: medial frontal gyrus, inferior prefrontal cortex, IPS in Thai and Chinese listeners</td>
</tr>
<tr>
<td>[61]</td>
<td>PET</td>
<td>same—different judgment of tones with short glides vs. long glides</td>
<td>LH: CB</td>
</tr>
<tr>
<td>[71]</td>
<td>IEP</td>
<td>passive listening to unvoiced vs. voiced CV syllables</td>
<td>RH: CB</td>
</tr>
<tr>
<td>[78]</td>
<td>fMRI</td>
<td>word identification based on differences in CLT vs. VOT</td>
<td>LH: HG and TP distinguish between voiced and unvoiced stops</td>
</tr>
<tr>
<td>[83]</td>
<td>DL</td>
<td>detecting the offset of monaural pulse sequence; delayed comparison of two durations</td>
<td>LH: IFG</td>
</tr>
<tr>
<td>[19]</td>
<td>DL</td>
<td>identification of CV syllables presented to the left and right ear simultaneously with 40- vs. 80-ms formant transition</td>
<td>RH: CB</td>
</tr>
</tbody>
</table>

fMRI = functional magnet resonance imaging; DL = dichotic listening; IEP = intracerebral evoked potentials; PET = positron emission tomography; EEG = electroencephalography; MEG = magnetencephalography; CB = cerebellum; STG = superior temporal gyrus; STS = superior temporal sulcus; MFG = middle frontal gyrus; IPS = inferior parietal sulcus; IFG = inferior frontal gyrus; IFS = inferior frontal sulcus; MTG = middle temporal gyrus; HG = Heschl’s gyrus; TP = temporal plane; SMA = supplementary motor area; BH = both hemispheres.
that using linguistically relevant stimuli (i.e., words) should result in left hemisphere lateralization.

Some inconsistencies exist with respect to the BG and the cerebellum contributing to speech timing. While some studies report an involvement of one or both structures, most of the studies cited in the preceding text do not. This is in many cases not surprising and can be explained by the comparisons that have been performed. If the BG and the cerebellum are involved in general timing operations that apply to non-speech as well as to speech stimuli, comparisons between non-speech timing and speech timing will cancel out significant effects in these regions. Moreover, as we know from mismatch negativity studies that listeners time acoustic events preattentively, it seems reasonable to assume that passive listening to hums that differ in duration also involves timing. Therefore, using passive listening as a baseline might ‘reduce’ activity in areas that are involved in general timing operations.

6. Comparing the timing of speech and non-speech events

Time is a fundamental dimension of behavior. Increasing interest in how we process temporal constraints in our environment and adjust our behavior accordingly has motivated many studies in the field of cognitive neuroscience. The main focus has been on interval timing—the ability to perceive and produce durations in the seconds to minutes range. Different models have been proposed that try to explain timing behavior. One of the most influential models is illustrated in Fig. 2. This model integrates different processing stages: the clock stage, where the passing of current time is monitored; the memory stage, where durations are stored that are of significance to the individual; and the decision stage, where current time is compared with stored time [46]. Research on interval timing aimed at investigating the brain structures that subserve processing at these different stages. Most of the techniques that have been employed require participants to compare the duration of an auditory or visual stimulus with the duration of a previously learned or upcoming stimulus. Given that these techniques cover durations that range from seconds to minutes, they shed light on the timing mechanisms underlying a broad range of human behavior. However, they fail to address the timing mechanisms inherent to speech perception and production, which operate on a finer temporal scale.

The present paper gave an overview of the studies that investigated the neuroanatomical substrates of timing speech. Although this research is still fragmentary, converging evidence from different studies highlights the role of certain cortical and subcortical structures in temporal processing. Moreover, this research suggests that left frontal and superior temporal cortex, the BG, and the cerebellum are critically involved in the timing of speech. In order to get some sense of what these structures might do, the following sections will integrate the different findings from language research and relate them to proposals that have been formulated on the basis of interval timing research. It will be interesting to see whether the mapping of brain structure and function that goes back to psychological models of interval timing can be applied to the timing of speech.

6.1. Cortex

Both patient and neuroimaging evidence suggests that the left cortical hemisphere is specialized for the processing of speech-relevant durations (e.g., Refs. [44,45,120]). While the patient literature that investigated global, syntactic, and lexical-semantic aspects of speech timing does not allow to distinguish between the role of different regions within the left hemisphere, patient studies investigating segmental aspects point toward a different role of anterior and posterior left hemisphere structures (e.g., Ref. [17]). Patients with anterior lesions do show more severe and qualitatively different deficits from posterior patients in distinguishing between stop consonants on a temporal scale. Interestingly, these deficits are restricted to the timing of consonants with the timing of vowels being

![Fig. 2. The processing model of interval timing as proposed by Gibbon et al. [46] incorporates a clock, a memory, and a decision stage. The pacemaker is thought to release paces. Passing time is monitored via the integration of these paces. The comparator decides whether passed time matches a behaviorally relevant duration. If this is the case a response is initiated.](image-url)
A specific role of left frontal cortex for the temporal processing of consonants has been confirmed in neuroimaging studies (e.g., Ref. [78]). Moreover, neuroimaging evidence supports a dissociation between temporal and frontal regions. In contrast to left temporal cortex, left frontal cortex is more strongly engaged in the processing of consonants, which contain rapidly changing formant transitions, as compared to steady-state vowels [33]. Additionally, left frontal regions were found to be more strongly involved in timing when the sounds that have to be timed carry some linguistic meaning [44,45].

A contribution of frontal structures to timing has also been implicated by interval timing research. Patients with frontal lesions have been found to underestimate durations [86] or to show deficits in time perception acuity [76]. As these studies pooled over participants with LHD and RHD, they do not allow conclusions about lateralization effects. Harrington et al. [54] investigated the lateralization of interval timing by looking at the performance of LHD and RHD patients separately. In contrast to research on speech, they found that temporal processing was specifically affected following damage to the right hemisphere. Interestingly, the temporal deficit of RHD patients was correlated with their ability to switch attention. Given that RHD patients with temporal deficits had lesions in either the middle and superior frontal gyri or the supramarginal and angular gyri, Harrington et al. [54] proposed a prefrontal-inferior parietal network for attention to time. They contrast this network with a right inferior-frontal superior-temporal pathway that assists in the retrieval and rehearsal of temporal information. It seems that this proposal could be applied to the timing of speech. However, the memory representations of linguistically relevant durations might be lateralized to the left rather than the right hemisphere. Moreover, the two hemispheres might differ in the attention that they attribute to certain types of temporal information. In contrast to the right hemisphere, the left hemisphere might direct attention preferably to stimuli with a fast rate of acoustic change.

6.2. Basal ganglia

Compared to the number of studies that investigated the role of the cortical hemispheres in timing speech, there is little research on subcortical structures such as the BG. Nevertheless, evidence from patient studies points to an involvement of the BG in temporal aspects of speech. The timing of speech production (e.g., number and length of pauses [34]) and speech perception (e.g., CLT [50]) is compromised in patients with BG dysfunction. Neuroimaging evidence for the involvement of the BG in speech timing is relatively sparse [100]. Moreover, the majority of studies that investigated temporal processing of speech failed to report activations in the BG [21,44,45,78]. However, as discussed earlier in the text, this might be explained by comparisons with a baseline task that also involved timing.

The interval timing literature considers the BG as one of the ‘hot spots’. A series of pharmacological studies in rats revealed evidence that the BG modulate the speed of an internal clock [79,80]. While the administration of dopamine agonists (e.g., methamphetamine) has been shown to increase clock speed, dopamine antagonists (e.g., haloperidol) were found to slow it down. Furthermore, patients with BG dysfunction tend to underestimate time ranges of a second and shorter [53,87]. Interestingly, a similar tendency has also been reported for the processing of time in speech [20,50]. Additional evidence for a role of the BG in interval timing comes from neuroimaging studies [57,98,99]. More specifically, activation of the BG, the thalamus, and the cortex has been reported for different tasks and modalities such as the timing of auditory and visual events [82] or the timing of finger movements [98].

Based on these findings, many researchers assume that the BG are involved in processes at the clock stage [47,52,77,81]. As part of a frontal-striatal circuitry, the BG are thought to contribute to timing by controlling an internal time keeper and by accumulating passing time that will be matched onto durations stored in memory. If these are more general mechanisms underlying every form of timing, they should also operate in the case of speech perception and production. Evidence from patients with BG dysfunction suggests that this might be the case. However, as this evidence is relatively sparse, this question should be tackled in further studies.

One possible way of testing whether the more general timing operations performed by the BG operate in the case of speech is to directly compare the timing of speech and non-speech events within the same group of PD patients. For example, PD patients who are trained to produce a short and a long duration will overestimate the short duration and underestimate the long duration in a later production test [75]. A comparable effect within the speech domain has been reported for the production of VOT in PD patients [70]. Similar to the migration effect in interval timing, the VOT distributions of voiced and unvoiced stop consonants have been found to overlap. Thus, it would be interesting to correlate the performance of BG patients with respect to both measures. Furthermore, it would be interesting to see whether patients that underestimate non-speech durations also underestimate durations in speech. To this end, one might compare a task where PD patients indicate whether a tone is closer in duration to a previously learned short or long standard tone with a task where they have to decide whether a CV syllable starts with a voiced or unvoiced stop consonant. Another aspect of interval timing that has been found compromised in BG dysfunction is scalar property [75]. Scalar property refers to the fact that uncertainty in timing is proportional to the duration being timed: small variance usually obtained for the timing of short durations will superimpose with the large variance obtained for the timing of long durations if the time axis is normalized with respect to the mean of each timing distribution. It seems reasonable to assume that the timing of speech events (e.g., voiced and unvoiced stops) also shows
scalar property. If true then a positive correlation between distorted scalar property for speech and non-speech timing in BG patients would provide further support for the idea that the BG are involved in more general aspects of timing that feed into the application of time in other cognitive domains such as speech.

6.3. Cerebellum

The cerebellum is a further structure that seems to contribute to certain aspects of timing speech. First, cerebellar disease affects the timing of consonant production, especially with increasing demands on articulatory coordination [2]. Second, perceptual temporal discrimination such as the distinction between voiced and unvoiced stops in word medial position is compromised in cerebellar patients [3]. Furthermore, there is neuroimaging evidence for a contribution of the cerebellum in speech timing tasks [78]. These findings have been attributed to an involvement of the cerebellum in temporal aspects of speech. Moreover, a complex interaction between the cerebellum and cortical regions implicated in speech, sensory-motor coordination, and working memory is thought to mediate timing. These cortico-cerebellar circuits are thought to be recruited in the case of high cognitive or motor coordination demands. Thus, the cerebellum may not be involved in the monitoring of passing time but in the application of temporal information in working memory or motor processes.

Similar assumptions have been formulated in the interval timing literature. For example, Gibbon et al. [47] propose that rather than being involved at the clock stage the cerebellum contributes to the decision process in timing tasks (see Fig. 2). Furthermore, Harrington and Haaland [52], although acknowledging the disruptive effect of cerebellar lesions on timing, draw attention to the cerebellum’s role in nontemporal processes and call for their dissociation from temporal processes in studies that investigate timing. Rao et al. [99] were able to dissociate nontemporal from temporal processes using fMRI. They asked participants to discriminate between two successive time intervals and recorded the BOLD signal at several time points between the onset of the first interval and the participants’ response. FMRI analysis revealed early activations in the BG that were associated with the temporal encoding of the two intervals. In contrast, the cerebellum was activated immediately before and during the response, which suggests an involvement in processes at the memory or decision stage rather than at the clock stage. Moreover, the authors propose that the cerebellum monitors and optimizes sensory input from the cortex thereby facilitating duration comparison in working memory.

6.4. Conclusions

Together, lesion and neuroimaging studies implicate an involvement of cortical and subcortical structures in the timing of speech. Comparing these structures with findings from interval timing research reveals some commonalities. Both speech and interval timing tasks recruit the BG and the cerebellum. With respect to cortical contributions, however, speech and interval timing elicit different patterns of lateralization. While the former recruits primarily left hemisphere structures, the latter is lateralized to the right hemisphere. Given these commonalities and differences it seems that speech involves both speech-specific and more general temporal processes. While cortical contributions might be speech specific in that they represent a specialization for durations that are critical for speech and the mapping of these durations onto linguistic meaning, the BG and the cerebellum might perform more general timing operations that feed into other cognitive processes such as the processes specific to speech. Several models of language processing and development already stress the idea that speech does not represent a function that evolved in isolation but that it builds upon other cognitive processes [25,117]. This paper suggests that timing is one of them and, thus, strongly encourages research that tries to abstract and correlate the processes that speech shares with other cognitive functions.

Acknowledgements

I would like to thank D. Van Lancker Sidtis, J.J. Sidtis, T.B. Penney, W.H. Meck, C.K. Friedrich, A.D. Friederici, Cindy Lustig, and an anonymous reviewer for constructive comments on earlier versions of the manuscript.

References


C. Liegeois-Chauvel, J.B. de Graaf, V. Laguitton, P. Chauvel, Specialization of left auditory cortex for speech perception in man depends on temporal coding, Cerebral Cortex 9 (1999) 484–496.


R. Näätänen, The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMN), Psychophysiology 38 (1999) 1–21.


G. Schaltenbrand, The effects on speech and language of stereotactical stimulation in the thalamus and corpus callosum, Brain Lang. 2 (1975) 70–77.


J. Schwartz, P. Tallal, Rate of acoustic change may underlie hemi-


