Brief article

Speech perception engages a general timer: Evidence from a divided attention word identification task

Laurence Casini [a],[*],[1], Boris Burle [a],[1], Noël Nguyen [b],[2]

[a] Aix-Marseille Université, CNRS, Laboratoire de Neurobiologie de la Cognition, Marseille, France
[b] Aix-Marseille Université, CNRS, Laboratoire Parole et Langage, Aix-en-Provence, France

ARTICLE INFO

ABSTRACT

Time is essential to speech. The duration of speech segments plays a critical role in the perceptual identification of these segments, and therefore in that of spoken words. Here, using a French word identification task, we show that vowels are perceived as shorter when attention is divided between two tasks, as compared to a single task control condition. This temporal underestimation pattern is consistent with attentional models of timing and hence demonstrates that vowel duration is explicitly estimated using a central general-purpose timer.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Speech unfolds over time. This has a crucial impact on how speech is perceived and understood. Among the many acoustic properties associated with speech segments, segmental duration itself is used by listeners over many different levels of processing. A well-known example relates to variation in vowel duration depending on whether the following consonant is voiced or voiceless. In many languages, vowels are shorter prior to a voiceless compared with a voiced obstruent (e.g. /e/ is shorter in "bet" than in "bed") (Kingston & Diehl, 1994). This context-dependent variation in vowel duration has been shown to influence the perception of post-vocalic consonant voicing: consonants are more often perceived as voiced rather than voiceless following longer vowels (Fischer & Ohde, 1990).

Despite the large amount of experimental evidence supporting a role for segment duration in speech perception and comprehension, current models of speech processing do not yet offer an accurate characterization of the mechanisms that allow listeners to deal with segment duration.

An important question is whether these mechanisms are specific to speech, or whether the perceptual handling of segment duration is accomplished by means of central timing mechanisms. Most speech perception models implicitly espouse the first of these two competing assumptions. In the now widespread exemplar-based models (see Nguyen, Wauquier, & Tuller, 2009, for an overview), for instance, speech sounds are mapped by the listener onto large sets of exemplars associated with words in long-term memory. Exemplars contain fine-grained information about the words' sound shapes, which, we can assume, includes segment duration. Such models lead to the suggestion that listeners' sensitivity to segment duration is both implicit and specific to speech, since it appears to be a by-product of the pattern-matching process and of the way in which exemplars are encoded in memory. The speech-specific approach to segment duration processing can, however, be challenged on empirical grounds. For example, studies on patients with basal ganglia dysfunction have shown that these patients, known to exhibit deficits in general temporal processing (Pastor, Artieda, Jahanshahi, & Obeso, 1992), are also impaired

* Corresponding author. Tel.: +33 4 88 57 68 78; fax: +33 4 88 57 68 72.
  E-mail address: Laurence.Casini@univ-provence.fr (L. Casini).
[1] Université de Provence, Laboratoire de Neurobiologie de la Cognition, Case C, 3 Place Victor Hugo, 13331 Marseille Cedex 03, France.
[2] Laboratoire Parole et Langage, 5 Avenue Pasteur, 13100 Aix-en-Provence, France.

when processing segment duration in speech (for a review, see Schirmer, 2004). This points to the possibility that the same general temporal mechanisms may be involved in both speech and non-speech temporal processing.

When the duration of sensory stimuli is explicitly processed, one of the most systematically observed performance patterns is that durations are perceived as shorter under conditions of divided attention (Brown, 1997; Casini & Macar, 1997; Coull, Vidal, Nazarian, & Macar, 2004; Hicks, Miller, Gaes, & Bierman, 1977; Macar, Grondin, & Casini, 1994; Thomas & Weaver, 1975). This underestimation has been attributed to the fact that time estimation is based on the number of "temporal pulses" accumulated over the measured interval: since some of these pulses are lost under attentional distraction, the duration of the target interval is judged to be shorter (Buhusi & Meck, 2005; Burle & Casini, 2001; Casini & Macar, 1997; Zakay, 1989). In this paper, we investigate whether attentional manipulations affect the perception of duration of speech segments in the same way. If so, it would suggest that processing of segment duration in speech is governed by a timer that is functionally similar to the one implicated in processing of non-speech duration.

When applied to context-dependent variation in vowel duration, this hypothesis predicts that vowels should be perceived as shorter under attentional distraction as compared to a control condition. This in turn should lead listeners to more frequently perceive the carrier word as ending in a voiceless, rather than a voiced, consonant.

In this experiment, we endeavored to provide empirical support for this proposal using a dual-task paradigm. In dual tasks, participants perform two tasks simultaneously. Therefore, they allocate less attention to each task as compared to a single task condition. This causes performance degradation in both tasks, as reflected by an increase in performance variability (Navon & Gopher, 1979; Sperling & Melchner, 1978). In addition, when the primary task requires temporal judgment, stimulus duration is perceived to be shorter than in the single task condition (for review, see Brown, 1997). We assessed this hypothesis using two French words, *cache* /ka∫/ "hiding place" and *cage* /kaʒ/ "cage", which end in a voiceless and voiced fricative, respectively. This choice of linguistic material was motivated by results obtained in two pilot studies. In the first of these studies, we recorded four French male speakers while they pronounced words ending with either a voiceless final fricative (*cache*, *gâche*, *mâche*, *hache*) or a voiced one (*cage*, *gage*, *mage*, *âge*). Each item was pronounced nine times by each speaker. In keeping with the fore-mentioned tendency observed for other languages, results showed that the vowel /a/ was significantly shorter in syllables with a voiceless coda (173 ms) compared with a voiced one (241 ms). In the second pilot study, we examined whether French listeners were sensitive to vowel duration when identifying the voicing of word-final consonants. To this aim, we constructed a series of five synthetic /kaC/ sequences in which the duration of the vowel ranged from 150 to 310 ms in 40-ms steps, and in which the final consonant was replaced by white noise (see Stimuli synthesis in the Methods section). Six participants performed a binary-choice (*cache* or *cage*) word identification task. The re-
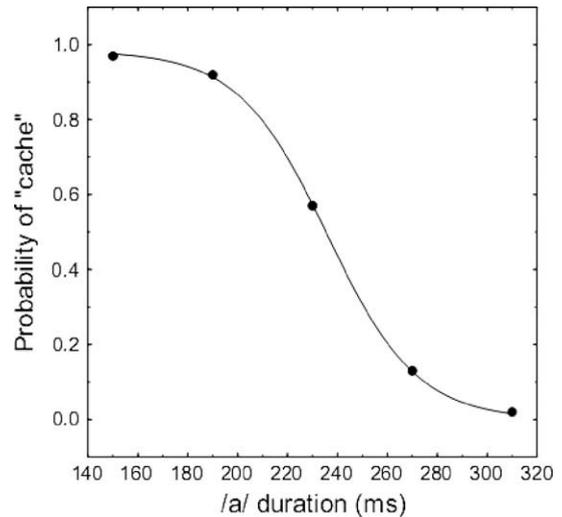
Fig. 1. Probability of "cache" responses as a function of /a/ duration (ms). A logistic function is fitted to the average probability value across the six participants.

sults showed that the proportion of words perceived as *cache* decreased as the duration of the vowel sound /a/ increased (Fig. 1), confirming that vowel duration is a perceptually-relevant cue to the voicing of the following consonant in French.

In the main experiment, eighteen French-speaking volunteers performed this word identification task either as a single task (as above) or at the same time as a secondary visual reaction time task (dual-task condition). We predicted that participants would identify the sequence as *cache* rather than *cage* more frequently in the dual task compared to the single task condition.

2. Methods

2.1. Participants

Eighteen participants took part in the experiment (nine women and nine men aged from 18 to 42 years). All participants were volunteers and gave informed consent to the experimental procedure, following the Helsinki declaration (World Medical Organization, 1996).

2.2. Procedure

The participants were seated in a dimly lit, sound-proof room. They faced a black panel (situated at a distance of 1 m) containing one light-emitting diode (LED) that lit up either red or green depending on the trial. Four response keys were available, two for each hand (index and middle fingers were used). The experiment was controlled by a microcomputer running t-scope (Stevens, Lammertyn, Verbruggen, & Vandierendonk, 2006).

There were three experimental conditions: a word identification task, a reaction time task and a dual task (word-identification + reaction-time task). After a training session, each participant completed two blocks of the word identification task, two blocks of the dual task and one

block of the reaction time task. The order of conditions was counterbalanced across participants.

2.2.1. Word identification task

The auditory stimuli were delivered to the participants through headphones and participants were required to indicate whether they perceived *cache* or *cage* by pressing the appropriate response key with the index or middle finger depending on the word.

The experimental block contained 110 trials corresponding to 11 different auditory stimuli, each delivered 10 times (inter-trial interval = 2 s).

2.2.2. Reaction time task

Participants had to press the response key, as quickly and accurately as possible, with the index or the middle finger depending on the color of the LED (red or green). The experimental block contained 110 trials (55 trials with red LED, 55 trials with green LED) with 2-sec inter-trial intervals. Reaction times measured the time between LED onset to the key press.

2.2.3. Dual task

The auditory stimulus was delivered simultaneously with the LED illumination. Participants were required to give the reaction time task response as quickly and accurately as possible, and then to provide the response to the word identification task. Thus, two responses were given for each trial, one with the right hand, and the other one with the left hand. The experimental block contained 110 trials corresponding to 11 auditory stimuli, each presented 10 times (five with green LED and five with red LED) with 4-sec inter-trial intervals.

The correspondence between the response hand and the task (reaction time task or word identification task) was counterbalanced between participants as was the correspondence between the finger used and the response provided (*cache* vs *cage* and *red* vs *green*).

2.3. Training session

2.3.1. Reaction time task training

The participants performed the reaction time task on twenty trials (10 green and 10 red stimuli, randomized order).

2.3.2. Word identification task training

We only used the shortest (/a/ duration of 150 ms) and longest (/a/ duration of 310 ms) stimuli. The participants were first presented with this pair of stimuli five times. No response was required. Next, they performed the word identification task with these two stimuli randomly presented ten times each.

2.3.3. Dual-task training

The participants performed the dual task on a randomized list of ten short (/a/ of 150 ms duration) and ten long (/a/ of 310 ms duration) stimuli.

The participants were first trained on the two single tasks (the order of which being counterbalanced across participants), then on the dual task.

2.4. Auditory stimuli synthesis

The auditory stimuli were synthesized using the HLSyn speech synthesis system (Sensimetrics). The HLSyn control parameters were derived from the word *cache* /kaʃ/ recorded beforehand by a French male speaker and submitted to a detailed acoustic analysis. The burst for /k/ occurred 15 ms prior to the vowel onset and was 10-ms long. The onset and target frequencies for the first formant (F1), the second formant (F2) and the third formant (F3) are presented in Table 1.

The duration of the onset-to-target interval was set to 40 ms. The fundamental frequency (F0) linearly decreased from 110 Hz at the beginning of the vowel to 90 Hz at the end of the vowel. A series of 5 (in the pilot perceptual experiment) or 11 (in the main experiment) stimuli was generated from this synthetic sequence. Each stimulus contained the sequence /ka/ immediately followed by white noise that replaced the final consonant (/ʃ/ or /ʒ/). Vowel durations varied from 150 to 310 ms in 40-ms steps (in the pilot experiment) or 16-ms steps (in the main experiment).

The words *cache* and *cage* both have a low lexical frequency in spoken French (respectively 3.88 and 16.61 in one million, as estimated in the Lexique 3 French lexical database, see New, Pallier, Brysbaert, & Ferrand, 2004). The number of CVC phonological neighbors ending in the same rime is close to being identical for the two words according to the Vocolex database (*cache*: 9, *cage*: 7, see Dufour, Peereman, Pallier, & Radeau, 2002).

3. Results

For each participant, we fitted a logistic function to individual performance in each condition, single or dual task, separately. This function allowed us to estimate two parameters (Fig. 2a): the standard deviation, reflecting perceptual variability, and the point of subjective equality (PSE) between the two words, reflecting the perceptual boundary between *cache* and *cage* along the vowel duration scale.

Under the dual-task condition, performance decreased in both tasks: reaction times were longer under the dual task (494 ms) than the single task condition (322 ms) (t_{17} = 5.44, p = 0.001) and, in the word identification task, standard deviation increased in the dual task (19.68 ms) compared to the single task (14.19 ms) (t_{17} = 2.5, p = 0.01).

More importantly, the mean PSE was located at a longer vowel duration value in the dual task (242 ms) compared to the single task condition (232 ms) (t_{17} = 2.4, p = 0.01) revealing a shift in the boundary between the two words (Fig. 2b).

Table 1
Onset and target frequencies for the first, the second and the third formant.

	F1	F2	F3
Onset frequency (Hz)	180	1970	2400
Target frequency (Hz)	620	1500	2360

*F*1 = the first formant, *F*2 = the second formant, *F*3 = the third formant.
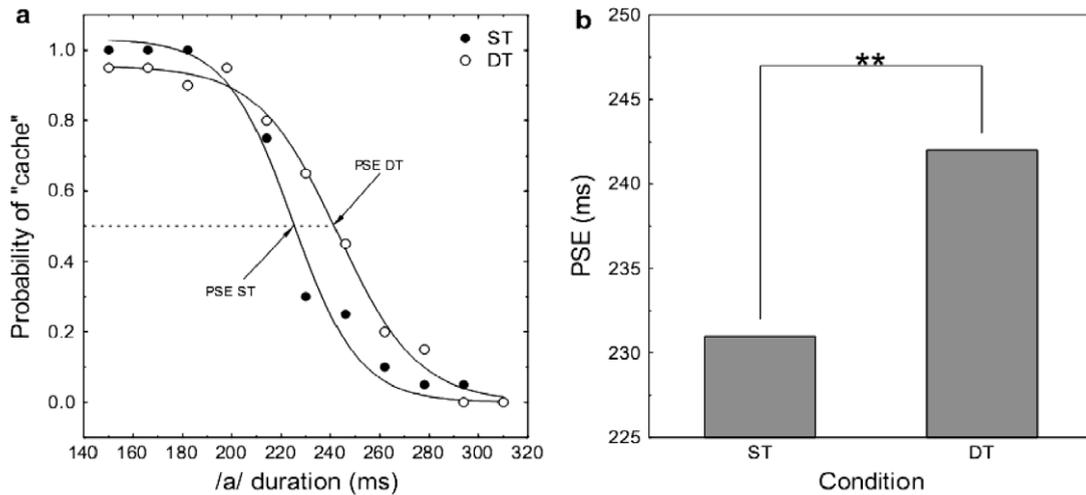
Fig. 2. (a) Probability of "cache" responses as a function of /a/ duration. A logistic function is fitted to the performance of one representative participant performing the word identification task in the single task (ST) and dual task (DT) conditions. PSE ST and PSE DT correspond to the point of subjective equality (PSE) estimated from the fit for the single- and dual-task conditions, respectively. Standard deviation corresponds to the tangent at the PSE. (b) Mean PSE (in ms) in the word identification task performed during the ST and DT conditions.

4. Discussion

Performing a dual-task procedure led to a decrease in performance of both tasks confirming that less attention was available for each task under dual-task conditions. More relevant to the focus of this experiment is the shift in the PSE towards longer vowel durations, in the word identification task performed under dual-task, compared to single-task, conditions. This shift indicates that perceptually ambiguous words were more frequently perceived as ending in a voiceless consonant (*cache*) in the dual task than in the single word identification task. These data show that the amount of available attentional resources affects vowel duration judgment and therefore has a significant influence on how words are recognized. More specifically, this finding suggests that divided attention alters word recognition because of an underestimation in segment duration. This raises the interesting question of whether, in continuous speech, words are recognized in a way which may vary depending on the listeners' level of attention: Can word identification be altered when the listener's attention is diverted away from the speaker?

Our results have important implications for studies on both time estimation and speech perception. First, they are at odds with the widespread assumption that the measurement of sub-second durations, such as those involved in speech, is accomplished by mechanisms distinct from those used for longer durations (for review, see Buhusi & Meck, 2005). In the case of sub-second durations, it has been proposed that measurement processes are automatic and derive from intrinsic properties of neural function whereas the measurement of longer durations would depend on a cognitively controlled timing system that uses attentional resources (Ivry & Schlerf, 2008). Our data provide evidence that is consistent with only few previous studies (Rammsayer & Ulrich, 2005; Thomas & Weaver, 1975) showing that sub-second durations are also sensi-

tive to attentional manipulation. In that respect, our results support the idea that the processing of sub- and supra-second durations share common mechanisms (Rammsayer & Ulrich, 2005). In addition, while attentional effects on explicit time estimation are well established, our results show that such effects also arise when, as in our experiment, the listener's response does not explicitly require a temporal judgment. Second, our findings demonstrate that estimation of speech segment duration involves a timer with the same sensitivity to attention as that implicated in duration estimation of non-speech stimuli. Models that have been proposed to account for these attentional effects in the literature so far, assume that there are cognitive mechanisms dedicated to the processing of duration that are independent of the nature of the stimulus (Casini & Macar, 1997; Hicks et al., 1977; Thomas & Weaver, 1975; Zakay, 1989; Zakay & Block, 1996). Our data therefore appear to provide evidence for the assumption that perceptual estimation of segmental duration in speech is governed by a general cognitively controlled timing system, as opposed to a speech-specific timer. This does not support the hypothesis of domain specific timers for different functions or modalities but rather favors the idea of a central amodal and non-specific timer.

The material used in the present study consisted of a single acoustic continuum, and this may make our findings open to an alternative interpretation. It may have been the case that the participants became aware of the association between vowel duration and final consonant identity, and so formed an explicit strategy of determining the duration of the stimulus and then responding with the associated word. It would then be this duration estimation strategy that would be subject to the influence of attentional load, consistent with the existence of a general timing mechanism as well as of a separate speech timing mechanism. We believe that the possibility for such a strategy to have formed was limited, however, because the participants

informally reported after the experiment that they had perceived the final consonant as being hidden by, as opposed to being replaced with, the white noise appended to the /ka/ sequence. This suggests that they did perform a word-recognition task rather than resort to an experimentally induced duration estimation strategy. Further work using a wider range of materials would, however, be needed to confirm this.

Although the neural substrates of time estimation are not yet clearly identified, basal ganglia (Harrington, Haaland, & Hermanowicz, 1998; Pastor et al., 1992) and cerebellum (Casini & Ivry, 1999; Ivry & Keele, 1989; Ivry, Spencer, Zelaznik, & Diedrichsen, 2002) have both been shown to be involved. This involvement also appears to be true for speech perception since patients with cerebellar lesions (Ackermann, Gräber, Hertrich, & Daum, 1997) or with basal ganglia dysfunction (Gräber, Hertrich, Daum, Spieker, & Ackermann, 2002) exhibit impairment in the estimation of segmental duration. Taken together, our findings, combined with those mentioned above, strongly suggest that far from relying on a speech-specific module, speech comprehension involves general timing processes that are also at work in many other cognitive activities. Integrating these processes into speech perception models now seems necessary.

Acknowledgments

References

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

Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception and Psychophysics, 59*(7), 1118–1140.

Burle, B., & Casini, L. (2001). Dissociation between activation and attention effects in time estimation: Implications for internal clock models. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 195–205.

Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience, 6*, 755–765.

Casini, L., & Macar, F. (1997). Effects of attention manipulation on perceived duration and intensity in the visual modality. *Memory and Cognition, 25*, 812–818.

Casini, L., & Ivry, R. B. (1999). Effects of divided attention on temporal processing in patients with lesions of the cerebellum or frontal lobe. *Neuropsychology, 13*, 10–21.

Coull, J. T., Vidal, F., Nazarian, B., & Macar, F. (2004). Functional anatomy of the attentional modulation of time estimation. *Science, 303*, 1506–1508.

Dufour, S., Peereman, R., Pallier, C., & Radeau, M. (2002). VOCOLEX: Une base de données lexicales sur les similarités phonologiques entre les mots français. *L'Année Psychologique, 102*, 725–746.

Fischer, R. M., & Ohde, R. N. (1990). Spectral and duration properties of front vowels as cues to final stop-consonant voicing. *Journal of the Acoustical Society of America, 88*, 1250–1259.

Gräber, S., Hertrich, I., Daum, I., Spieker, S., & Ackermann, H. (2002). Speech perception deficits in Parkinson's disease: Underestimation of time intervals compromises identification of durational phonetic contrasts. *Brain Language, 82*, 65–74.

Harrington, D. L., Haaland, K. Y., & Hermanowicz, N. (1998). Temporal processing in the basal ganglia. *Neuropsychology, 12*, 3–12.

Hicks, R. E., Miller, G., Gaes, G., & Bierman, K. (1977). Concurrent processing demands and the experience of time-in-passing. *American Journal of Psychology, 90*, 431–446.

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

Ivry, R. B., & Schlerf, E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences, 12*, 273–280.

Ivry, R. B., Spencer, R., Zelaznik, H., & Diedrichsen, J. (2002). The cerebellum and event timing. *Annals of the New York Academy of Sciences, 978*, 302–317.

Kingston, J., & Diehl, R. (1994). Phonetic knowledge. *Language, 70*, 419–454.

Macar, F., Grondin, S., & Casini, L. (1994). Controlled attention sharing influences time estimation. *Memory and Cognition, 22*, 673–686.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86*, 214–255.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, and Computers, 36*, 516–524.

Nguyen, N., Wauquier, S., & Tuller, B. (2009). The dynamical approach to speech perception: From fine phonetic detail to abstract phonological categories. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to phonological complexity.* Berlin: Mouton de Gruyter.

Pastor, M. A., Artieda, J., Jahanshahi, M., & Obeso, J. A. (1992). Time estimation and reproduction is abnormal in Parkinson's disease. *Brain, 115*, 211–225.

Rammsayer, T., & Ulrich, R. (2005). No evidence for qualitative differences in the processing of short and long temporal intervals. *Acta Psychologica, 120*, 141–171.

Schirmer, A. (2004). Timing speech: A review of lesion and neuroimaging findings. *Cognitive Brain Research, 21*, 269–287.

Sperling, G., & Melchner, M. J. (1978). The attention operating characteristics: Examples from visual search. *Science, 202*, 315–318.

Stevens, M., Lammertyn, J., Verbruggen, F., & Vandierendonk, A. (2006). Tscope: A C library for programming cognitive experiments on the MS Windows platform. *Behavior Research Methods, 38*, 280–286.

Thomas, E. A. C., & Weaver, W. B. (1975). Cognitive processing and time perception. *Perception and Psychophysics, 17*, 363–367.

World Medical Organization (1996). Declaration of Helsinki. *British Medical Journal, 313*(7070), 1448–1449.

Zakay, D. (1989). Subjective and attentional resource allocation: An integrated model of time estimation. In I. Levin & D. Zakay (Eds.), *Time and human cognition* (pp. 365–397). Amsterdam: North-Holland.

Zakay, D., & Block, R. (1996). The role of attention in time estimation processes. In M. A. Pastor & J. Artieda (Eds.), *Time, internal clocks and movement* (pp. 143–164). Amsterdam: Elsevier.