Tick Talk

Douglas J. Davidson (dvdsn@casper.beckman.uiuc.edu)
Kathryn Bock (kbock@s.psych.uiuc.edu)
David E. Irwin (dirwin@s.psych.uiuc.edu)
Beckman Institute
University of Illinois
405 N. Mathews
Urbana, IL 61801 USA

Abstract

Three experiments on the production of time expressions are reported. In each experiment, speakers produced time expressions to analog clock faces while their eye movements were recorded with a headband-mounted eye tracker. Previous work has shown that speakers can apprehend the visual-conceptual information in a clock display very quickly, and that this is followed by relatively slow, incremental formulation of the time expressions. Analyses of fixation patterns from the present experiments lend additional support to this model.

Introduction

Clock time-telling offers an unusually efficient and effective means of studying how speakers formulate linguistic expressions to relatively simple visual displays. American English speakers can produce either absolute (e.g., *nine thirty*) or relative (e.g., *half past nine*) time forms for the same clock, and this provides a contrast between two opposite word orders for the same visual display. Clock faces also contain a reasonably large response set. Considering just the times at five minute intervals (e.g., *9:30, 9:35, 9:40*...) an analog clock face contains a small combinatorial explosion of at least 276 unique time expressions, each corresponding to exactly the same hour and minute hand arranged in different orientations. In addition, there is an unambiguous mapping from the hands of a clock display to the components of time expressions. This mapping allows techniques such as eye tracking to be applied to study the relationship between gaze patterns and speech.

Eye tracking has emerged as a powerful tool to investigate language production because, among other things, it can be used to observe whether speakers seek visual information corresponding to the individual terms of a multi-word expression before they begin to articulate those terms. Under some circumstances, the properties of the recorded eye movements (e.g., gaze position, duration, sequence) can be related to the properties of the recorded speech (timing, expression choice, fluency) to provide evidence for or against hypotheses about the mechanisms of speech planning.

Griffin and Bock (2001) proposed a general decomposition of the process of formulating descriptions based on data from experiments in which participants produced spontaneous descriptions of simple visual events. The first phase in this account, apprehension, involves the rapid extraction of visual features and conceptual relationships relevant to the upcoming utterance. The second phase, linguistic formulation, involves retrieving words and constructing the expression form to convey the intended description. Importantly, linguistic formulation occurs more slowly than apprehension, and it is accompanied by gaze patterns that match the order of the spoken words.

Time telling provides a strong test of this hypothesized decomposition because unlike the description of a novel scene using a novel sentence, both clock reading and time telling are well-rehearsed procedures that many speakers perform habitually throughout the day. In addition, time expressions form a finite set. In principle, speakers could map each conceptual representation of a time to a geometrical representation of a corresponding clock in memory. This, in turn, could permit a fast mapping from the perception of a clock display to the formulation of the time expression, potentially bypassing slow formulation requirements.

Bock, Irwin, Davidson & Levelt (2003) showed that this bypass is not taken. In their experiments, speakers started to look to regions of the clock faces that corresponded to the first term of the ultimate expression that they produced approximately 400 ms after the onset of the clock displays, suggesting very fast apprehension for the clock displays. The signature pattern of successive gazes to regions corresponding to successive terms of utterances was as robust for clock displays as in earlier studies of scene description. The experiments reported here extend the results of Bock *et al.* by investigating several alternative explanations for the data patterns that were originally observed.

The speakers in Bock *et al.* were asked to produce time expressions as either absolute or relative time forms. By knowing the format required for the time expression in advance, in principle speakers could have adopted a task set that biased them to individuate the clock hands in the order of the required expression. For example, if speakers are aware that they should produce a time consisting of first the hour and then the minute value, an efficient strategy would be to first look for the hour hand. The task set, in this case, would be the primary determinant of the observed eye movement patterns, rather than an intrinsic tendency to link gaze patterns with speech planning processes. Left to their own devices, without explicit instructions to produce one or the other form, speakers could provide eye movement patterns that are less closely linked to planning processes. In Experiment 1 reported below, speakers were simply asked to produce time expressions without specific instructions about the required form. In addition, half of the participants heard instructions that presented a relative time expression as an example utterance, while the other half heard an absolute example.

The hands of analog clock displays are differentially informative, insofar as the hour hand conveys information about both the hour and the minute. This is the case because the hour hand moves parametrically from one hour to the next for each increment of the minute (see the top right clock in Figure 1). For example, one can deduce that the time is *9:30* just by observing that the hour hand is halfway between *9* and *10* –the hour hand could provide all the information necessary for the clock time. The hands are differentially informative because while the hour hand conveys this information about the value of the minute, the minute hand does not convey any information about the hour. The location of the minute hand at the *6* position, for example, does not provide the exact angular position of the hour hand. It only conveys that the hour hand is midway between some (unspecified) pair of hours.

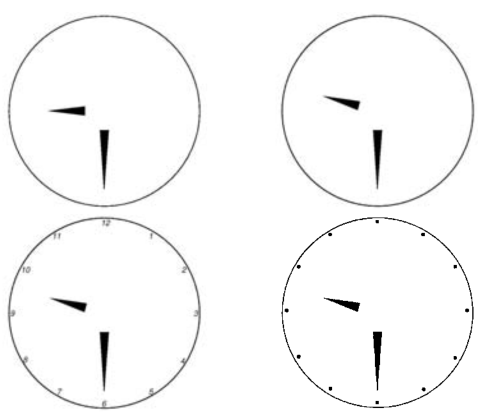

Figure 1: Numberless fixed, numberless parametric, numbered, and tick-marked clocks.

The analog clock displays used in Bock *et al.* did not have parametrically-arranged hour hands. In their stimuli, the hour hand remained fixed at one location for varying locations of the minute hand (see top left in Figure 1). It is possible that Bock *et al.* observed gaze paths that mimicked word order patterns because unlike a regular clock, the hour hand region of the clock face did not contain all of the information usually found in a clock display. Speakers might have *needed* to inspect both hands in those experiments to understand the time, whereas on a regular clock, the differential relationship of the hands could in principle permit speakers' gaze to favor the hour more. Experiment 1 investigates this possibility.

Among the requirements of time telling, a speaker must understand both *where* the hands are located, and to *what* the hands refer. On a clock face without any numbers, speakers must determine the angular relationship between the hand locations and some reference point, and they must determine which hour and minute regions are relevant.

Numbered clock faces may make this process simpler, because the digits present on an ordinary numbered clock face provide both identity and location information. The value for the hour can be acquired by simply reading the digit to which the hour hand points, and the value of the minute can be acquired by calculating the appropriate minute value after the digit has been read, or by retrieving this value from memory once the digit has been identified. Digits also encode location information by providing a point of reference for nearby hands, relative to the orientation of the entire clock face. It could be that either identity and location information is enough to circumvent the need for incremental gaze patterns during time phrase production. Experiments 2 and 3 compare numberless clock faces like those in Experiment 1 with numbered and tick-marked clock faces (see Figure 1) to examine the contribution of these factors to time telling.

Experiment 1

Method

Participants Sixteen participants from the Psychology Department subject pool at the University of Illinois took part for course credit or for monetary payment. All participants had normal or corrected to normal vision, and spoke American English as their native language. None had taken part in the other experiments reported here.

Design and Materials The experiment was a two-factor design with type of clock (fixed hour hand, parametric hour hand) and instruction (absolute biased, relative biased) as between-participant factors. Each participant saw 144 clock faces, consisting of all combinations of hour and 5-minute times. The fixed hour hand clock faces were constructed so that the hour hand always pointed to the location of the hour, while the parametric clock faces were constructed so that the hour hand advanced for each 5-minute increment of the minute hand.

Apparatus Eye movements were recorded using an headband-mounted Eyelink (SR Research, Ltd.) eye tracker. Clock displays were presented on a 51 cm monitor, each clock subtending approximately 23 deg of visual angle. Participants pressed a button on a hand-held button box to begin and end each trial. Speech was recorded using a sound card and a preamplified microphone.

Procedure Participants were simply asked to produce time expressions to the clock faces. One example of an absolute expression was provided in the absolute-biased instructions, and one example of a relative expression was provided for the relative-biased instructions. Participants were not given explicit instructions about the form of the time expressions.

Each trial began with the presentation of a central fixation point. Participants pressed a button to start a trial, and drift correction was applied while they held fixation to the center point. If drift correction was successful and participants held fixation, a clock face was presented for a maximum of five seconds, or until participants pressed a button indicating that they had finished the trial. The clock face remained on screen for a minimum of three seconds. The next trial began after a one second delay.

Data Analysis Participants' fluent utterances were classified as *absolute*, *relative*, or *full*-hour time expressions. Full hour time expressions were excluded from further analysis, because there was no choice of construction for those cases. Also, times corresponding to displays in which the number and minute hand overlapped (e.g., *2:10*) were excluded, as these did not contain different fixation target regions for hours and minutes.

For the fixation duration data, an extension of the standard Cox proportional hazards model was used to compare cumulative hazard rate functions of fixation durations in different experimental conditions. This extension allows for the analysis of repeated measurements and random effects in survival analysis (Therneau & Grambsch, 2000; also known as "frailty" models).

Results and Discussion

Speech Table 1 shows the average proportion of relative time expressions observed in all conditions. Speakers preferred to produce absolute expressions for both types of clocks and for both types of instruction by a wide margin. Given this preference, speakers were more likely to produce a relative expression when the instructions contained a relative example ($p < .001$), but there was no difference between the two clock types.

Table 1: Average Proportion of Relative Expressions.

	Clock Type		
Instruction Bias	Orig.	Param.	Ave.
Relative	0.113	0.103	0.108
Absolute	0.000	0.032	0.016
Ave.	0.056	0.068	

Further inspection of the data revealed that most of the relatives were produced by just two speakers. When they produced relative expressions, these speakers were more likely to do so for times near the top of the hour. An analysis of the proportion of expressions that were relative showed that there was a linear main effect of minute value ($p < .001$), as well as a quadratic trend ($p < .001$), indicating that these participants produced a few relatives for minute values 0-15, almost no relatives for times near the half hour, and progressively more relatives for minute values greater than 40. See Bock, Irwin, and Davidson (in press) for an account of this pattern.

There were no statistically significant differences for speech onset times for the speakers who saw the fixed clocks (mean latency: 1658 ms) versus the speakers who saw the parametric clocks (1769 ms).

Eye Movements Figure 2 shows a global profile of the gaze patterns for absolute expressions. It plots the average proportion of fixations to the hour and minute hands relative to the onset of the clock display (top plot) and to the onset of speech (bottom) for absolute expressions. The proportions were calculated by summing the number of fixations in 100 ms intervals over subjects.

The top plot shows that the two types of clock faces led to similar distributions of fixations to the hour and minute hands. Speakers first fixated the hour hand region and then the minute, regardless of whether the clock faces were fixed or parametric. The bottom plot of Figure 2 shows that, relative to speech, the peak in the proportion of fixations to the hour hand occurred earlier than the peak of the minute hand, and that the peak of the minute hand occurred near the onset of speech. Both the hour and minute distributions appear to be shifted somewhat earlier for the parametric clocks, which could be due to differences between the two subject groups.

There was a median of four fixations before the onset of speech across the experimental conditions. Table 2 shows the pattern of saccade transitions for the first four fixations, providing a more fine-grained analysis of the data shown in Figure 2. Each cell contains the proportion of transitions to the hour, minute, and other regions for each transition pair

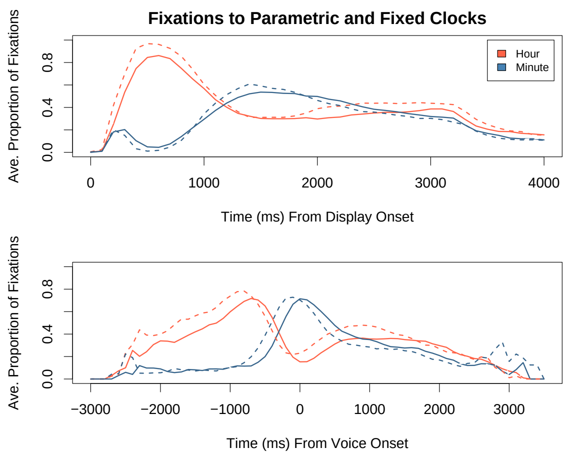

Figure 2: Average proportions of fixations to fixed (solid) and parametric (dotted) clock faces relative to display onset (top) and speech onset (bottom).

(transitions to the center region are not shown). Table 2 shows that the early fixations 1-4 were more likely to target the hour hand region. For example, at the transition from the first to the second fixation after leaving the center, 14.6% of transitions were from the minute to the hour, but there were fewer of the reverse, hour-to-minute transitions, 1.4%. This initial fixation to the hour hand was then most often followed by a series of refixations to the hour hand. Later, at the transition from fixation 4 to 5, there were more transitions to the minute region.

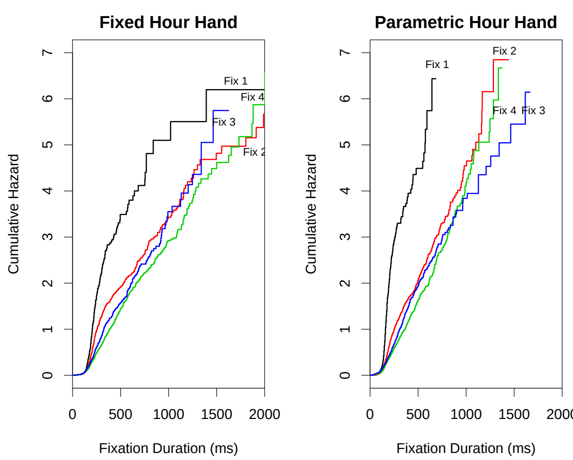

Figure 3: Cumulative hazard functions for the first four fixations of the fixed and parametric clocks.

Figure 3 plots cumulative hazard functions for fixations 1-4 after participants' gaze left the center region. These fixations correspond to fixations 1-4 in Table 2. The slope of the cumulative hazard function at a particular time point can be interpreted as the rate at which the fixations end in a saccade at that time point. Figure 3 shows that the first fixation after clock onset did not last as long as the fixations that came later, as the estimated cumulative hazard function for the first fixation has a steeper slope (exp(coef)=0.36, χ^2=1077, $p < 0.005$) than the later fixations, which appear to have similar, more shallow slopes. In addition, both the fixed and parametric hour hand clocks show this pattern.

The data from Experiment 1 suggest that speakers produced time expressions to fixed analog clock faces much like they do for parametrically-arranged faces. A similar global profile of fixations was observed for both types of clocks, and speakers did not appear to have more (or less) trouble with fixed clocks. The initial fixation after clock onset appeared to be shorter than the other fixations for both the fixed and parametric clocks.

Table 2: Transition matrices for fixations 1-4 (source region along the left side, destination region along the top).

Transition: Center:1

	Hour	Minute	Other
Center	0.436	0.215	0.311

Transition: 1:2

	Hour	Minute	Other
Hour	0.409	0.014	0.014
Minute	0.146	0.013	0.054
Other	0.139	0.021	0.150

Transition: 2:3

	Hour	Minute	Other
Hour	0.519	0.129	0.059
Minute	0.033	0.017	0.009
Other	0.036	0.056	0.135

Transition: 3:4

	Hour	Minute	Other
Hour	0.266	0.218	0.105
Minute	0.022	0.143	0.033
Other	0.020	0.079	0.103

Transition: 4:5

	Hour	Minute	Other
Hour	0.133	0.115	0.060
Minute	0.056	0.309	0.075
Other	0.038	0.077	0.117

Experiment 2

Method

Experiment 2 compared the performance of speakers to numbered clock faces versus numberless clock faces. In most respects except the design and stimuli, the experiment was like Experiment 1 (differences are outlined below).

Participants Eight participants from the University of Illinois community took part for course credit or pay. All participants were native speakers of American English, and none of the participants had taken part in the other experiments reported here.

Design and Materials Experiment 2 was a one-factor within-participant design with clock type (numbered, numberless) as the main experimental factor. The numberless clocks were the same stimuli as the parametric condition of Experiment 1 (see Figure 1, top right). The numbered clocks were again the same clocks, but with digits placed at the periphery (see Figure 1, bottom left). Each participant saw 144 trials total, consisting of 72 numbered and 72 numberless clocks. The stimulus lists for presenting the clocks were counterbalanced so that four of the participants saw one version of a clock as a numbered clock, while the other four saw the same clock as numberless (and vice versa). Instructions to participants were like those in Experiment 1, except that both relative and absolute time expressions were provided as examples (participants were told that either form would be an acceptable response).

Results and Discussion

Like Experiment 1, two speakers produced the majority of relative expressions in Experiment 2. Overall, a similar proportion of relatives (0.08 percent of the expressions) were produced as in Experiment 1, and they were more common near the top of the hour.

For the absolute expressions, speakers were faster to start time expressions for the numbered clocks (1638 ms) compared to the numberless clocks (2326 ms), $t(7) = 8.47, p < 0.001$, as expected. There were too few relative expressions across subjects to make a meaningful comparison.

Figure 4 shows the aggregate profile of fixations to the hour and minute hands over time for numberless and numbered clocks for absolute expressions. The data pattern suggests that speakers spent less gaze time at the time-relevant regions for the numbered clocks. The top plot of Figure 4 shows that after display onset, fixations to the hour hand started to decrease earlier, and fixations to the minute hand increase earlier for the numbered clocked compared to the numberless clocks.

The plot of fixations relative to speech (bottom plot) shows that the span between the peak of the hour hand region fixations and voice onset was somewhat shorter for numbered clocks compared to numberless clocks.

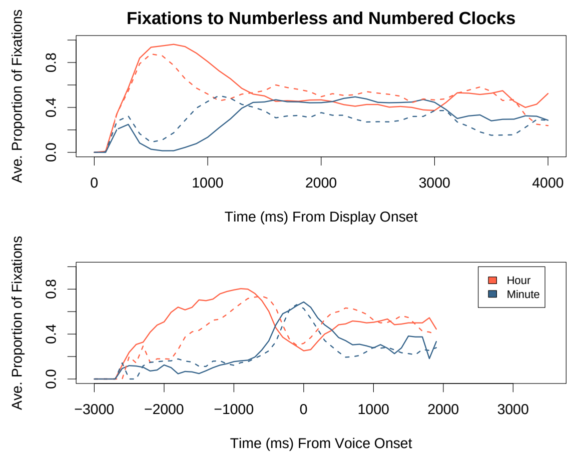

Figure 4: Average proportions of fixations to numberless (solid) and numbered (dotted) clock faces relative to display onset (top) and speech onset (bottom).

The data from Experiment 2 suggest that the display of number information accelerates the production of the hour term, but does not change the incremental pattern of formulation. The global profile of fixations showed that speakers were faster to leave the hour region to start fixations to the minute region for numbered clocks. Nonetheless, the same basic pattern of incremental looking first to the hour hand and then to the minute hand in concert with speech was observed for both types of clocks.

Experiment 3

Experiment 3 examined performance on clock faces without tick-marks to clock faces with tick-marks. In most respects except the design and stimuli, the experiment was like Experiment 2. The differences are outlined below.

Method

Participants Eight participants from the University of Illinois community took part in the experiment for course credit or pay. All participants were native speakers of American English, and none had participated in the other experiments reported here.

Design and Materials Experiment 3 was a one-factor within-participant design with clock type (tick-free, tick-marked) as the main experimental factor. The tick-free clocks were the same stimuli as the parametric condition of Experiment 1 (see Figure 1, top right), while the tick-marked clocks were the same clocks as the tick-free, but with small dots placed at the periphery (see Figure 1, bottom right). Both types of clocks had parametric hour hands. Each participant saw 144 trials total, consisting of 72 tick-free and 72 tick-marked clocks. The stimulus lists for presenting the clocks were counterbalanced as in Experiment 2, and the instructions were the same.

Results and Discussion

In Experiment 3, one participant produced the majority of relative expressions (57), while two others produced 15 or less (out of the 132 opportunities for each subject). The majority of the relative expressions had minute values at 45, 50, and 55, as in the other experiments.

For the absolute expressions, speakers were faster to produce times to the tick-marked clocks (1764 ms) compared to the tick-free clocks (2004 ms), $t(7) = 4.17, p < 0.005$. There were too few relatives across subjects for a meaningful comparison.

Figure 5 shows the aggregate proportion of fixations to hour and minute hands for clocks with and without tick-marks. Like Experiment 2 with numbered clocks, it appears that speakers were somewhat faster to leave the hour hand and start fixating the minute hand with the tick-marked clocks, although the effect is smaller than with the numbered clocks. There appears to be little effect of the tick-marks for the span between fixations to the hour region and the onset of speech.

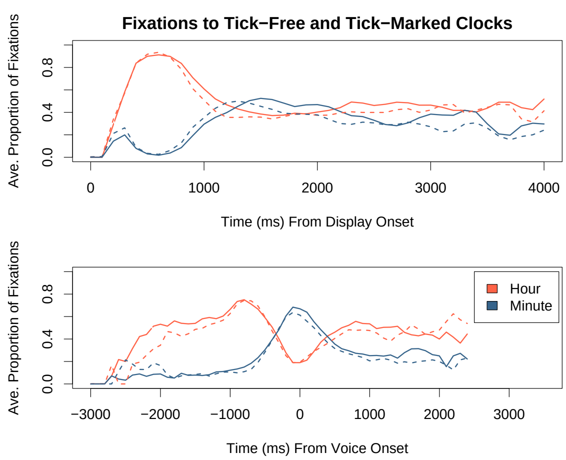

Figure 5: Average proportions of fixations to tick-free (solid) and tick-marked (dotted) clock faces relative to display onset (top) and speech onset (bottom).

Experiment 3 shows that the tick marks do not greatly affect the patterns of fixations or performance of time telling to the clock faces. The same incremental pattern of fixations was observed, although shorter latencies were observed for tick-marked clocks.

General Discussion

It appears that speakers' gaze patterns are largely insensitive to small changes in the details of analog clock faces, while at the same time, speech performance parameters may be more strongly affected. Across all three manipulations, most speakers chose to produce absolute time expressions, and they provided very similar fixation patterns. When numbers were present on the clock faces, speakers were faster to start utterances and appeared to be faster to move from one hand to the next, but otherwise provided similar fixation profiles. If speakers had faster access to phonological information by reading the numbers in Experiment 2, this may have enabled them to spend less time gazing at the relevant time regions. These patterns would be consistent with results from Meyer and van der Meulen (2000) suggesting that a substantial factor influencing fixation times during object naming is phonological form retrieval. The effect of tick-marks on the fixation profiles was small, but did affect utterance latencies as well.

In general, the data from these experiments provide additional support for hypothesized properties of formulation in language production, and for the usefulness of eyetracking and measures of eye-voice coordination during time telling in assessing those properties. Across all three experiments, when producing absolute expressions, speakers appeared to locate the hour hand region quickly (within the first or second fixation). This was followed by a series of fixations that reflected the order of the words that participants produced. It appears that even with highly formulaic expressions and a fixed and finite vocabulary, the eyes still foreshadow the path of speech.

Acknowledgments

This research was supported by research and training grants from the National Science Foundation (SBR 94-11627 and SBR 98-73450) and the National Institute of Mental Health (RO1-MH66089 and T32-MH18990).

References

Bock, J. K., Irwin, D., Davidson, D. J. & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language*, *48*, 653-685.

Bock, J. K., Irwin, D. E., & Davidson, D. J. (in press). Putting first things first. J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: What we can learn from free-viewing eyetracking*. New York: Psychology Press.

Griffin, Z. & Bock, J. K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274-279.

Meyer, A. & van der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review*, *7*, 314-319.

Therneau, T. M. & Grambsch, P. M. (2000). *Modeling survival data: Extending the Cox model*. New York: Springer-Verlag.