The Effects of a Speeding World: Auditory Perception and Recall of Information at Differing Words per Minute

Luke K. Brown

Indiana University

Abstract

There's been a large amount of research pertaining to the number of words per minute one can understand. However, there appears to be a gap in connecting understanding to recall and even more important, performance. This concern grows ever more major as we move towards substantial integration with technology. Since we are constantly looking to be more efficient with our time, this paper explores whether there is a tradeoff between playback speed and ability to recall and utilize the incoming information. This tradeoff was explored by having subjects view three different videos of the same category at different speeds (1x, 1.5x, and 2x) and then testing them on the knowledge they learned. Although the results were somewhat inconclusive, words per minute had no correlation to the number of correct answers (up to 2x speed). In future studies, it'd be interesting to determine where this cutoff of comprehensible speech and ability to retain semantic information lies.

The Effects of a Speeding World: Auditory Perception and Recall of Information at Differing Words per Minute

As technology progresses we have the option to experience an excessive amount of incoming information. Whether it's using the word search function in order to find the proper paragraph within a PDF, listening to a podcast at two times speed, or even using artificial intelligence to plan our daily tasks, our demand for efficiency is ever increasing. In this study, I wanted to observe the effects of this increased rate of incoming information; specifically, in regards to playback speed and the effects it has on memory recall in order to record what was the most efficient speed that still allowed one to retain information.

In addition, as we have shifted to a mixture of online learning as a result of the COVID-19 pandemic, this topic becomes more and more important. It is especially important for classes with an asynchronous style of learning where it's up to the students discretion as to how fast they desire the lecture to play back. The question arises, will this increase in incoming information affect one's ability to capture and retain what is heard for use on a quiz, exam, or future life application?

Literature Review

Over the years, there have been several proposed theories on how humans perceive language. Although my research doesn't necessarily pertain to hearing recognition models, I believe these models are still relevant in understanding the processes and underlying mechanisms. In addition, the research shows the speed at which particular levels of these models occur. With this, we are able to extrapolate a possible amount of comprehendible words per minute.

One of the first and most prominent of these models was William Marslen-Wilson and Paul Warren's COHORT model (Raynor & Clifton, 2008). Wilson's first model only had a serial bottom-up form of processing, but with later research he altered the model to also include top-down parallel processing (Raynor & Clifton, 2008). The COHORT model has three distinct phases: access, selection, and integration (Marslen-Wilson & Warren, 1994). The access stage occurs once the very first phoneme has been uttered (Marslen-Wilson & Warren, 1994). For instance, if I were to say "att__", the listener would access their lexicon of words and determine a multitude of possibilities: "attach", "attention", "attack", "attitude", etc. In the selection phase, the number of possible words is narrowed down, or "selected" based upon the rest of the phonemes from the spoken word. In regards to selection, if I were to say "London", the listener would be able to select the word as soon as 'd' was uttered as there are no other words within their possible lexicon that would fit the progression of the pronunciation. The integration portion checks whether the semantic and syntactic meaning of the selected word aligns within the context of the spoken sentence. These three phases make up the bottom-up serial portion of the model. The parallel processing comes into play as the listener is able to discern the appropriate word faster in the access phase based upon the context of the words around it within the spoken sentence (Marlsen-Wilson & Warren, 1994).

Marslen-Wilson and Warren's data shows that from start to finish (access to integration) the processing for a particular one to two syllable word is 200 ms (Marslen-Wilson & Warren, 1994). This can be converted to seconds and becomes 0.2 which approximates to about 300 selections (words) per minute. A similar model, named TRACE, created by James McClelland and Jeffrey Elman reduces that number to about 185 ms which gives an estimated number of 324 selections per minute (McClelland & Elman, 1986).

With past and current research, there appears to be a similar convergence on speaking rate at 120-200 words per minute; however, there also appears to be a significant amount of disagreement in the number of words one can comprehend per minute as Raynor and Clifton Jr. describe that 340 to 400 words per minute are able to be heard (Raynor & Clifton, 2008). Heingartner's article agrees with that and adds to it by stating that an individual can understand words beyond the 400 word per minute mark (Heingartner, 2003). Heingartner even reports that a blind man named Gregory Rosmaita can go upwards of 650 words per minute; however, there may be increased audio perception abilities caused by his lack of visual experience. Heingartner explains that as the rate of words per minute increases, there might be a correlation between how much a listener actively focuses on the incoming words which leaves people worse off to recall information playing at slower speeds than faster speeds (Heingartner, 2003). At any rate, it is agreed upon that the listening rate is substantially faster than the words per minute one can speak.

In this experiment, my aim was to observe if there was a tradeoff between information absorbed and the speed at which a video was playing. As such, I predict that 2x speed will yield the worst performance, followed by 1x, which leaves 1.5x for the optimal speed to percent correct ratio.

Methods

For this experiment, I chose three different videos all relating to abstract philosophical explanations: Existentialism, Utilitarianism, and Contractarianism. The similarity in concept is extremely important since this study has a within-subject design. If the videos were of substantially different topics (art, history, science, etc.) there's a possibility that one of the subjects would be more well-versed in one of those areas which would introduce a confound

causing their performance at different speeds to be contingent upon the subject matter, not the speed. Each video is on Crash Course's YouTube page with the same host (Hank Green) giving the information to the audience. The consistency with the speaker is extraordinarily important as a different speaker can affect the talking speed which in turn affects the words per minute.

In regards to the participants I chose twelve individuals and gave them the speed at which to play the video (A, B, or C) back at. I made them report the speed in the form in order to be sure that the required rate was chosen. Those three speeds (1x, 1.5x, and 2x) were selected as they are available to all YouTube users without the requirement of an additional plugin. In order to eliminate a learning bias as well as potential differences within each video, four permutations were given (A, B, C… C, A, B… B, A, C… B, C, A). Out of the twelve participants there were three groupings (four subjects per "group") with corresponding speeds per each video. The speed within each of these permutations was consistent per each grouping:

Group 1: A → 1x, B → 1.5x, C → 2x with permutations per each individual in the group

Group 2: A → 2x, B → 1x, C → 1.5x with permutations per each individual in the group

Group 3: A → 1.5x, B → 2x, C → 1x with permutations per each individual in the group

As stated previously, this is a within-subject design, so each participant is compared to themself and the groupings are only to ensure that each video gets a desired speed (as the words per minute at each speed are similar). The dependent variable is the number correct out of ten and the independent variable is the playback speed (or more precisely, the words per minute). Words per minute is more precise as each video has a slight difference in overall words per minute, and speed (1.5x, 2x, etc.) doesn't account for these very slight differences.

There were ten questions associated with each video that the participants were to answer immediately following the video. The questions were all multiple choice and pertained to

concepts and explanations found within each individual video (nothing arbitrary like what shirt was the host wearing). I also asked that the subjects would turn off closed captions in order for there to be greater equality between all subjects. Each quiz was created inside of Google Forms. The first, second, and third page was the same for every form: general instructions, subject informational data, and the link to the video with instructions on how to change the playback speed. The fourth page was where the ten questions were asked.

After the data was collected, I performed a variety of statistical tests within R. These tests determine the significance of speed, video, and the correlation between words per minute and number correct.

Results

Table 1: Preliminary word per minute data of each video (and lectures for a baseline)

Video	Words	Minutes	wpm	wpm*1.5	wpm*2	Crash Course : C343 Lecture	Crash Course : Q370 Lecture
A	1879	8.883333	211.5196	317.2795	423.0393	1.531641	1.527218
B	1969	10	196.9	295.35	393.8	1.425778	1.421660
C	1885	9.516666	198.0735	297.1103	396.1471	1.434276	1.430133
C343 lecture 18 (20:38-30:38)	1381	10	138.1	207.15	276.2		
Q370 lecture Oct 23 (10:00-20:00)	1385	10	138.5	207.75	277		

*Note: For each lecture a 10 minute portion of the video was taken and then compared to the wpm (words per minute) of all three videos (A, B, and C) giving a ratio of how much faster the Crash Course videos were than a "normal" talking speed

Table 2: ANOVA for analysis of main effect of speed and video on number correct

Effect	DFn	DFd	F	P	Significance ($p < 0.2$)
Speed	2	22	0.1587629	0.8541674	

Video	2	22	9.577947	0.00101835	*

*Note: No interaction can be said about video and speed as each individual had a separate speed per each video

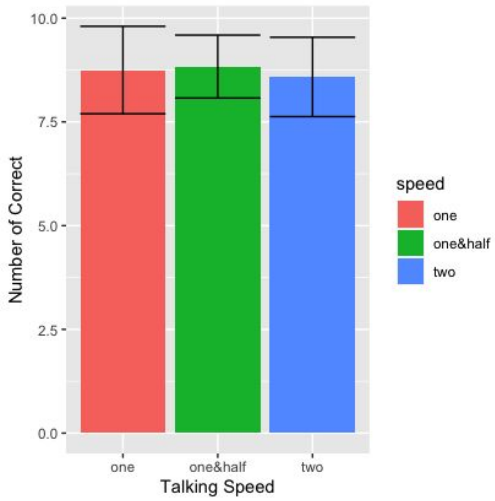

Figure 1: The relationship between the number correct and talking speed

Table 3: Mean number correct per each speed

1x	1.5x	2x
8.750000	8.833333	8.583333

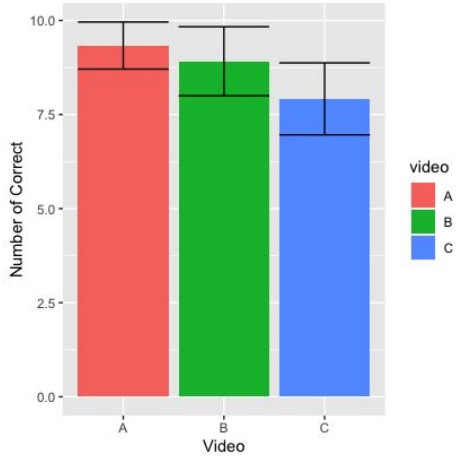

Figure 2: The relationship between video and number correct

Table 4: Mean number correct per each video

A	B	C
9.333333	8.916667	7.916667

Table 5: Mean per each video with the corresponding speed

	A	B	C
1x	10	8.50	7.75
1.5x	9	9.50	8.00
2x	9	8.75	8.00

*Note: Since no one was given the same video twice, this is comparing the number of correct for all subjects per each video at a designated time (groups of 4 individuals per each video at a determined speed as mentioned in the methods). To make this more clear I have designated 2 out of the 4 groups with associated colors (green being the overlap).

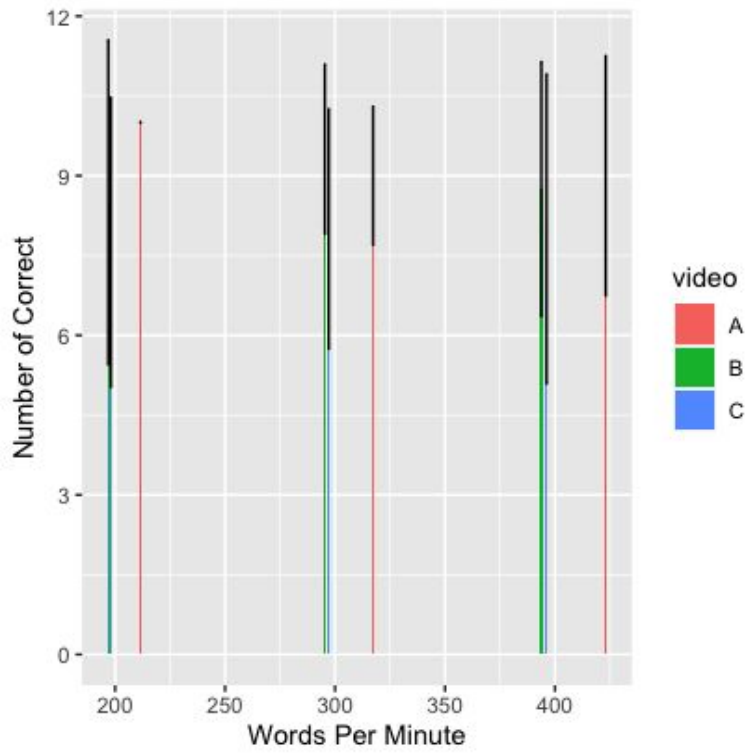

Figure 3: The relationship of number correct with the associated words per minute of each video

Table 6: Words per minute of each video with the associated number correct

196.9 (B - 1x)	198.073 5545 (C - 1x)	211.519 6998 (A - 1x)	295.35 (B - 1.5 x)	297.110 3317 (C - 1.5x)	317.279 5497 (A - 1.5x)	393.8 (B - 2x)	396.147 1089 (C - 2x)	423.039 3996 (A - 2x)
8.50	7.75	10.00	9.50	8.00	9.00	8.75	8.00	9.00

Table 7: Correlation between number correct and words per minute

t	df	p	Significance ($p < 0.2$)
-0.11321	34	0.9105	

Discussion

In reference to Table 1 from above, the typical talking speed for lectures was taken in order to obtain a baseline for the wpm (words per minute) that occur in a "normal setting". These values were obtained by accessing the transcripts of the associated video and then dividing the number of words based upon the length of the clip. This value was then compared to the videos used in this experiment (A, B, and C), and it was found that the Crash Course videos were already 1.5x faster than the baseline number of wpm. This is not an issue; however, it is interesting as one would be able to increase the speed on the lectures substantially more than on the Crash Course videos to obtain around the same wpm. This also points to certain times of speed being inconsistent in overall words per minute, so when addressing speeds it's always important to refer back to the wpm. With this being said, let's look at Table 2, the significance of the dependent variables.

In this study I have found there is not a significant main effect in the loss of semantics depending on the rate of incoming information (up to 2x the normal speed) $F(2, 22) = 0.1587629$,

$p > 0.2$. On the same note, Table 7 shows there is also no correlation between the number correct and wpm $r(34) = -0.11$, $p = 0.91$. Looking at Figure 1 and Table 3 we can see the visualization in the lack of change between the speeds with 1.5x having the highest number of correct answers (8.83) followed by 1x (8.75) and then 2x (8.58). Even though the result wasn't significant, my hypothesis was still correct that the scores for 1.5x would be the best, followed by 1x, leaving 2x with the worst number of correct responses (although not by much). Like stated in the literature review (and this is just mere speculation as there is no significance) but the small discrepancy could possibly be attributed to the speed of 1x not being engaging enough to capture our complete attention. As the speed increases there's a chance that one actively pays more attention until there is a plateau in the number of words one can understand in a minute. However, this plateau was never reached in this study since we were only dealing with the default YouTube playback options. With this being said, Table 6 shows that the highest wpm (423.04) ties for third place out of all the contenders. However there might be a cause for concern due to discrepancies *between* the scores on the videos.

There was a significant main effect in regards to the videos $F(2, 22) = 9.577947$, $p < 0.2$. Looking at Figure 2 and Table 4 we can see that people performed best on video A (9.33 - also the video with the highest possible words per minute), followed by B (8.92), leaving C in last place (8.58). The disparities between the videos can be due to a variety of factors. The one that comes to mind first is the difficulty of questions asked post-video. However, I asked the participants later on and they said that they were all very similar in difficulty (obviously this is word of mouth and the data doesn't show what they claim). It could also be the overall concept of the video. I attempted to pick arbitrary philosophical videos, but maybe Existentialism (video A) and Utilitarianism (video B) are more common and familiar than Contractarianism (video C)

which would lead to poorer performance on video C compared to the others. Another possibility is that more intelligent people were given video C at 2x speed than 1x speed leading to an overall lower performance on C as 2x speed would (potentially) be harder to follow than 1x. However, one thing that contradicts this theory is the score for 1.5x was still around both the score for 1x as well as 2x (one would expect it to be significantly higher if my previous theory was correct). I'd like to refer one's attention back to Table 5 and Figure 3. The data highlighted in yellow was the first three subjects I ran in order to test this very potential confound before I began to test other participants. As you can see with this highlighted data alone, performance became worse as speed increased. However, video C was played back at 2x speed for this grouping (with two out of getting a near perfect). So what appeared promising to begin with actually showed to be an issue later on. However, one thing that's certain is a difference in wpm between the videos did not cause the discrepancy since the words per minute of C and B were almost identical, and Table 5 shows there was a substantial difference in the amount of correct answers between the different speeds of those two videos.

Even though I found that speed was insignificant (up to 2x) and there was no correlation between wpm and score, my data still agrees with Raynor, Clifton Jr., and Heingartner's range of understandable wpm which refutes the numbers given by both Marlsen-Wilson and Warren as well as McClelland and Elman. In order to improve upon the study it would be very beneficial to run more individuals per each speed grouping in order to eliminate the possibility of an intelligence confound. In addition, one should very thoroughly make sure that both the topics per each video as well as the questions wouldn't lead to a difference in performance (which would be difficult due to the subjectivity surrounding difficulty). I also think it would be interesting to determine where the cutoff between wpm and performance would lie. At any rate, I think it's

safe to say that one can enjoy their videos up to 2x speed without any loss of semantic information.

References

Heingartner, D. (2003, October 2). *Now Hear This, Quickly*. https://www.nytimes.com /2003/10/02/technology/now-hear-this-quickly.html.

Marslen-Wilson, W. D., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, *101*(4), 653–675. https://doi.org/https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1037%2F0033-295X.101.4.653

McClelland, J. L., & Elman, J. L. (1986). The TRACE Model of Speech Perception . *Cognitive Psychology*, *18*, 1–86.

Rayner, K., & Clifton, C. (2008). Language Processing in Reading and Speech Perception is Fast and Incremental: Implications for Event Related Potential Research. *Biological Psychology*, *80*(1). https://doi.org/https://dx.doi.org/10.1016%2Fj.biopsycho.2008.05.002