Laws for Pauses

David L. Gilden and Taylor M. Mezaraups

University of Texas at Austin

Author Note

David L. Gilden, Department of Psychology, University of Texas at Austin

Taylor M. Mezaraups, Department of Psychology, University of Texas at Austin.

D. Gilden developed the study concept and study design. Testing and data collection were the primary responsibility of T. Mezaraups. Both authors conducted the data analyses. D. Gilden drafted the manuscript with editorial advice from T. Mezaraups. Both authors approved the final version of the manuscript. We thank Neil Doughty for conducting a pilot version of the first study as part of his Honors Thesis under the supervision of the first author. We also thank Alyse Kelly and Zöe Feygin for assistance in running the second experiment. Spreadsheets containing the raw data analyzed in Experiments 1 and 2 are located at the Texas Data Repository:

https://dataverse.tdl.org/dataset.xhtml?persistentId=doi:10.18738/T8/YEMGNA

https://dataverse.tdl.org/dataset.xhtml?persistentId=doi:10.18738/T8/6WBIY5

Correspondence concerning this article should be addressed to David L. Gilden, Department of Psychology, University of Texas at Austin, 108 E. Dean Keeton Stop A8000, Austin, Texas 78712. E-mail: dgilden@utexas.edu

Abstract

It is shown that a particular class of pauses taken in both read and composed speech obey allometric laws such that mean pause length predicts body size. The pauses in this class have durations that roughly span 250 ms to 1,000 ms and are taken to mark grammatical and prosodic boundaries. A theory of pause allometry is developed based on the observation that these pauses are expressive, they give speech momentum and rhythm, and most importantly, their durations reflect temporal discrimination – they are not produced by articulatory constraints. The theory is formulated in terms of a leaky integrator differential equation that is intended to model the sense of time passage that occurs during relatively brief pauses. The theory predicts that if the decay time scale associated with the leakage term includes body size as a parameter, then allometry will be observed generally in the amount of silence people deploy in pause behavior. A second study tested the theory on a class of long pauses defined by being terminated by a speech gesture indicating speaker recognition that the pause was indeed long. These long pauses were also found to obey allometry. The exponents derived from power law models of mean pause duration in both studies were found to be significantly larger than those associated with allometries of body energy expenditure. These findings provide a new meaning to the embodiment of cognition.

Keywords: allometry, memory, speech, language, ethology

Laws for Pauses

Ethology offers a window into human temporality through the simple expedient of cataloguing how time is spent in various forms of behavior. In this article we catalogue time spent in just one type of activity, that of pausing in the course of normal speech, and demonstrate that pause durations, on average, satisfy allometric laws. As allometry is not a common mode of analysis in psychological research, some clarification of how it is distinguished from correlational analysis may be useful. Allometry begins with the observation that much of animal morphology, physiology, and behavior is correlated with the single trait of body size. The study of these correlations forms the field of allometry, first conceptualized in terms of development and the differential growth of body parts relative to overall body size (ontogenetic allometry), but later generalized to include both physiological and behavioral characteristics of adult animals (static allometry). Allometric laws are generally framed as power laws of mass such that *animal property* = a*mass^{b}. The animal properties that have this form of size scaling are diverse and include, for example, head size (anatomy), heart beat period (physiology), burst acceleration (behavior), and so on. Although all allometries are based on statistical correlations, the focus of allometry is generally not on testing the hypothesis that a non-zero correlation between body size and an animal property exists. Rather, allometry is viewed as a form of measurement that produces a scaling exponent, *b*. The focus on *b* derives from its importance as a constraint on biological theory.

As preface to the studies, the methodological problems that have been encountered in the ethology of action duration are discussed, and it is argued that speech pauses are optimal for a close study of human temporality. A brief discussion of speech pause phenomenology follows that introduces the class of medium length pauses, between 250 and 1,000 ms, investigated in the

first study. The data that supports the demonstration of allometry in this class is then presented, and it will be clear that the relevant effects are both large and general. An analytic theory of how allometry might arise from the way people use their sense of time passage to terminate pauses is developed. This theory has generality beyond pauses that occur in fluid speech and is tested in a second study on a class of long pauses, 750 to 1500 ms, that are taken while people are planning answers to questions. These pauses are of sufficient size that they are terminated by a filled pause, "um" or "uh", to indicate that an answer is forthcoming. It is shown that these long planning pauses, while not occurring as part of fluid speech, also display allometry.

Issues in the Ethological Measurement of Consumed Time

The multi-cultural investigation of ordinary behavior conducted by Schleidt and colleagues (Schleidt, Eibl-Eibesfeldt, & Pöppel, 1987; Schleidt, 1988; Schleidt & Feldhütter, 1989; Feldhütter, Schleidt, & Eibl-Eibesfeldt, 1990) illustrates, albeit indirectly, the inherent complexity of an ethological approach to timing. The principal finding from this work is that the elementary action units that comprise typical human activity have durations universally concentrated about a single characteristic value of two to three seconds – a value associated with the span of the subjective present (Pöppel, 1997). However, from a methodological point of view, the more instructive point may be that in none of this work was there a clear description of how elementary action units were identified, suggesting that the methods were largely informal (see White, 2017).

An ethology of action that distills fluid behavior spanning minutes into elementary action units spanning seconds confronts the difficult problem of inferring beginnings and endings from the motions of body parts. There is considerable subtlety to this because an elementary action unit is not a single ballistic motion of a body part, but rather a sustained pattern of motion that

coheres into a meaningful gesture. Action units formed from motion repetition, such as in combing, cutting, nodding agreement etc., are particularly relevant to working, grooming, and social behavior, and are also particularly resistant to rule based protocols for segmentation. Repeated motions may continue for many tens of seconds with intermittent pauses and body part adjustments, and this choreography cannot be resolved into discrete action units without deciding which pauses and which adjustments are segmenting. These decisions have considerable impact on study outcomes because slight variations in the rules for segmentation can lead to large differences in the distributions of action unit duration. An instructive example of this sensitivity comes from ethological investigations of chew bursts, the action unit associated with mastication. The measured mean duration of chew bursts ranges from 3 sec (Gerstner & Cianfarani, 1997) to 13 sec (Po et al., 2011), depending on whether pauses of respectively 1.5 or 2 sec are regarded as marking endings. As there is no theory that sets the exact size of the watershed pause, there is also no basis for a principled construction of chew burst duration distributions. Exactly the same problem bedevils attempts to form distributions of speech burst durations on the basis of a critical pause length (Kien & Kemp, 1994). In fact, any ethology that attempts to identify action units on the scale of seconds will have to contend with the exquisite sensitivity that unit sizes have to small perturbations of segmentation rules.

A rigorous study of how time is spent in ordinary behavior requires the identification of a domain where 1) the rules for identifying the units of analysis may be formalized, and 2) where duration distributions of these units are not greatly sensitive to rule implementation. Silent speech pauses are, in this regard, optimal. A decibel cut-off on an acoustic waveform suffices to operationalize pause onsets and offsets, meaning that pauses can be extracted by a rote algorithm. Moreover, slight perturbations to the decibel cut-off will lead to proportionally slight

perturbations in the pause distributions. As a consequence, the distributions of speech pause durations are easily and meaningfully formed, and this circumstance has led to a substantial literature on pause duration phenomenology.

A Very Brief Introduction to Speech Pauses

Pauses sort into distinguishable classes on the basis of their duration. At the brief end of the continuum, 100 – 200 ms, *articulatory pause*s appear both within and between words and reflect either physical limitations of the organs involved (Dalton & Hardcastle, 1977) or serve to facilitate the perceptual interpretation of speech. The word "happy", for example, contains an articulatory pause at the double consonant. Zellner (1994, Fig. 1) illustrates the waveform produced by a speaker who inserts a 100 ms pause at that juncture. *Segmenting pauses,* as they will be referred to here (also referred to variously as hesitation or syntactic pauses), form a separate class with durations in the approximate range of 200 to 1,000 ms. They are placed at prosodic and grammatical boundaries (Yang, 2004, Krivokapic´, 2007) and effectively set the overall rate of normal speech (Goldman-Eisler, 1968). The purposes served by this class of pauses are often framed in terms of the practical aspects of speech production such as speech planning and speech recovery (Krivokapic´, 2007), but much of this article is concerned with the subtle role that these pauses play in creating speech cadence. And finally, people may stop speaking for a second or more for any number of reasons – turn taking, pondering, disfluency, and so on. This informal classification of pause duration is reflected in bimodal models of articulatory pauses and segmenting pauses in read speech (Campione & Veronis, 2002; Demol, Verhelst, & Verhoeve, 2007), and in a trimodal model of spontaneous speech that also includes a class of long pauses centered at about 1.5 s (Campione & Veronis, 2002).

A Gestalt Framework for Thinking about Pause Duration

In contrast to the durations of the elementary action units identified by Schleidt and colleagues, the durations of the pauses of interest here, segmenting pauses, do not organize around a characteristic value within a community of speakers. In part, this is due to the circumstance, that within individual speakers, segmenting pause durations are dependent upon the speech context in which they occur (see Ferreira, 1991, 1993; Goldman-Eisler, 1958, 1968; Krivokapić, 2007). Consequently, measures of central tendency of segmenting pause duration distributions do not reflect a unitary construct, such as the span of the present moment, but rather reflect the collective action of many factors that influence pause duration. Simply put, the factors that influence pause duration are apparently more complex and various than the factors that set the durations of elementary action units - peeling, hammering, scratching, etc. The absence of a single characteristic value for pause duration in individuals does not, however, preclude the possibility that there is organization at the level of the community. That is, different people may have different potentialities or properties that influence their pause behavior in systematic ways. Here we consider body size as a property that may be active in influencing pause duration. In order to motivate this idea, it will be helpful to conceptualize the functions of pauses in terms that go beyond linguistics and which involve central constructs in Gestalt psychology.

Temporal organization becomes an issue in pause phenomenology through the semantic content and sentence shaping that pauses may confer. Through the manipulation of silence, a speaker can, for example, indicate what they think is important, interesting, odd, imperative, and so on. Pause structure also manifestly gives speech abstract properties of shape such as rhythm and cadence. In the language of Gestalt, cadence and the semantics of silence are emergent

properties that arise from the perceptual organization of the speech signal. Although the introduction of a construct such as emergence may threaten the rigor that is typically brought to the quantitative analyses of pauses, it is both appropriate and potentially significant in view of our earlier work where we found that allometric laws had a central role to play in temporal organization (Gilden & Mezaraups, 2021).

Gilden and Mezaraups (2021) investigated how time delays interposed between successive events affected the temporal integration of those events into a common scene or group. Specifically, we were interested in characterizing the maximum pause duration that might be interposed before temporal grouping was disrupted. This maximum pause duration acts as a temporal proximity constraint for that grouping process, and it is the case that, for many forms of grouping, the proximity constraint evaluates at 2±1 s [Footnote 1]. Gilden and Mezaraups presented a theory of why proximity constraints are generic to temporal organization, and this theory led to the conjecture that the entire class of proximity constraints would satisfy allometric laws. The conjecture received support through an empirical analysis of proximity constraints in the experience of rhythmic pulse and in the perception of illusory paths in apparent motion. In these contexts allometry in a proximity constraint means that body size predicts both the slowest tempo that a person can feel rhythm and the slowest rate of image alternation that allows two successive images to be fused into a single moving object. Recognizing that the durations characteristic of segmenting pauses are too short to express proximity constraints in the grouping of words into speech phrases, the fact that segmenting pauses are intimately involved in the creation of emergent speech rhythms suggested that they might also obey allometric laws. As pause durations are readily extractable from speech, this question was not difficult to settle.

Experiment 1: Allometry in the Pause Durations in Read and Composed Speech

In this study we investigated whether the durations of segmenting pauses satisfy allometric laws. Pauses with duration shorter than 250 ms were excluded from the study, to ensure that articulatory pauses did not enter the sample. This limit will be justified below. There was no strict upper limit imposed on the pause duration distribution but, in our speech tasks, there were very few pause durations that exceeded 1,000 ms; this acts as an effective upper bound for our sample. Five different speech tasks that included examples of both composed and read speech were investigated. In this phase of investigation, we were principally interested in the existence, robustness, and generality of allometric scaling relationships.

Method

Participants

Sixty-eight native English speakers were recruited from the undergraduate subject pool at the University of Texas at Austin and received course credit for their participation. Ages ranged from 18 to 25 years, and heights ranged from 59 to 77 inches tall, with a median of 66 inches.

Stimuli

The study consisted of a series of five speech tasks that were chosen to represent a range of ordinary language behavior. Participants' speech acts were recorded using a Samson Meteor Mic USB Studio Condenser Microphone and Audacity open-source digital recording software. A description of each task follows: *Question/Answer* - two questions were presented that the participant was instructed to answer in detail. The first question asked for the participant's college major and why they chose it, and the second asked for their favorite animal and why they like it best. These questions were inevitably followed by sufficiently lengthy replies that a pause duration analysis could be meaningfully conducted. *Map task* - a map of a city grid was

presented, and the participant was instructed to give clear and detailed directions from the starting to end points. Depicted on this map were a series of arrows that indicated the desired path, which meandered through several streets and landmarks, requiring a relatively detailed set of directions. *Picture task* - the grayscale cartoon line drawing from the Bransford Balloon Study (Bransford & Johnson, 1972) was presented, and participants were asked to explain in detail what was depicted. This cartoon is both interesting and unusual, and so prompted relatively lengthy descriptions. *Poem task* - the beginning of the poem "One Fish, Two Fish, Red Fish, Blue Fish*"* by Dr. Seuss (1960) up to and including "Funny things are everywhere" was presented, and participants were prompted to read the poem clearly. This specific poem was chosen for its relatively simple rhythmic structure, its familiarity, and for its being fun. *Paragraph task* - participants read the memory stimulus from the Bransford Balloon Study that described the cartoon used in our picture task. This specific paragraph was chosen because it is written in a neutral sounding voice and is highly readable.

Procedure

Participants were physically situated to ensure a relatively constant recording level across tasks, in order to simplify the extraction of pauses from the audio signal. The experimenter explained to participants that they would be completing five speech production tasks and that their voices would be recorded. It was emphasized that there were no right or wrong answers. The tasks were then presented in a random order, except for the paragraph task, which always followed the picture task. Each task was displayed on the screen individually, and the experimenter first read the instructions to the participant and answered any questions they had. After it was clear that the participant understood the task, the read or composed speech prompted by the task instructions was recorded.

Pause Extraction

Analysis of the speech records into pauses and speech bursts was accomplished by employing Audacity's Sound Finder tool on the audio waveforms. A decibel level of -28 dB within Audacity, corresponding to 4% of the maximum resolvable signal, was set as the threshold of silence. Transition points in the waveform, where the signal level crossed this threshold at both speech offsets and speech onsets, were eligible for receiving a marker. The choice of a -28 dB threshold reflected our experience that thresholds greater than -28 dB often cut off the ends of words, marking them as silence, while thresholds less than -28 tended to misrepresent periods of ambient noise during silence as speech.

In order to isolate segmenting pauses and exclude articulatory pauses, it was necessary to set the minimum pause duration (marker separation) for which markers would be placed. Thresholds employed for this purpose are typically set between 200 and 300 ms (Hieke, Kowal, & O'Connell, 1983), but values as low as 100 ms are also not uncommon (Goldman-Eisler, 1958; Henderson, Goldman-Eisler, & Skarbek, 1966; Butcher, 1981). Goldman-Eisler (1968) employed a cut off at 250 ms, arguing that while such a choice may result in the loss of some very short grammatical and prosodic pauses, it does guarantee a purer sample of segmenting pauses by effectively excluding articulatory pauses. We followed this rationale in setting the minimum pause length at 250 ms. So configured, the Sound Finder Tool produced marker placements that matched in detail with those that we placed by hand in preliminary testing.

An example of how the audio track is partitioned by the Sounder Finder tool is shown in Fig. 1. This particular recording is from the poem task, and the words contained in the audio file

are written above the signal at their initiation points.

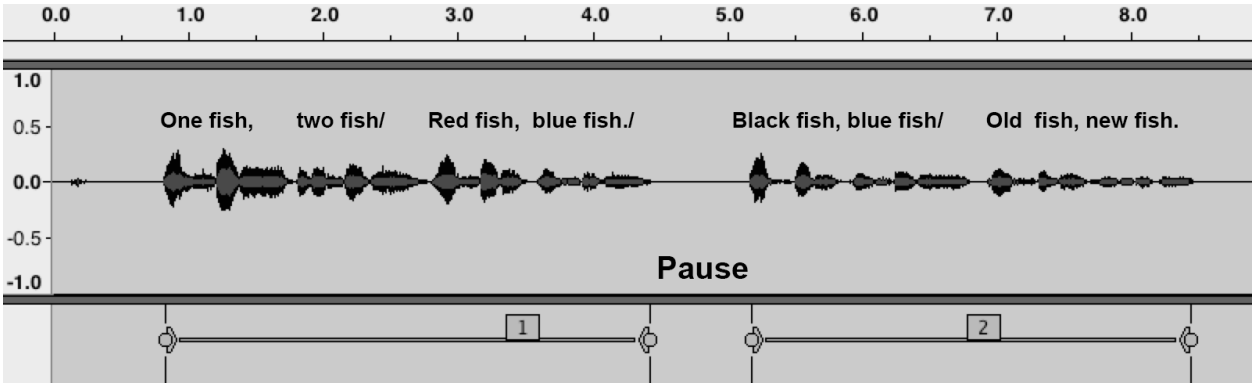

Figure 1. Screenshot of a speech waveform in Audacity with accompanying words from the poem task. Also shown are labels marking the speech burst terminals and the extraction of a pause. The pause duration is computed as the latency between the ending of burst 1 and the onset of burst 2.

Fig. 1 depicts two speech bursts separated by a single pause. The speech bursts are marked below the waveform with the labels 1 and 2, and the pause is the period of silence between them. This particular reader does not respect line endings (after "two fish" and "blue fish") but does take a significant pause at the middle of the fish enumeration. This pause occurs at a place in the text marked by a period, but this period is not the end of a well-formed sentence. Rather, the pause and the period mark a natural resting spot encouraged by the poetic structure. A more dramatic reader might have paused for line endings and commas, but this was not the case here and also not generally among our sample of undergraduates. Also evident in the waveform are pauses of about 200 ms associated with the stop consonants /k/ in "black" and /d/ in "old. In this study, we specifically wished to exclude pauses produced by the mechanics of articulation and only to include pauses produced by prosodic and grammatical landmarks. By

setting the minimum pause length to 250 ms, the segmenting pause between the two fish phrases is labelled, while the shorter articulatory pauses are not.

In practice, there were occasions when the Audacity pause marking algorithm would produce an errant result. Errors might be caused by vowels or entire words falling below threshold and getting counted as pauses, room sounds splitting pauses, and word endings being truncated as the amplitude falls below threshold. The waveforms that trigger these sorts of errors are depicted in Fig. 2 along with the corrective actions taken. The first panel depicts a word spoken below the assigned threshold, triggering a pause event; the corrective action is to delete the pause. The second panel depicts a percussive room noise occurring during a pause that causes the pause to split in two; the corrective action is to restore the original pause as if the room noise had not occurred. The third panel shows how a pause may be extended by a word ending falling below threshold; the corrective action is to shorten the pause to reflect the moment when the word in fact ended.

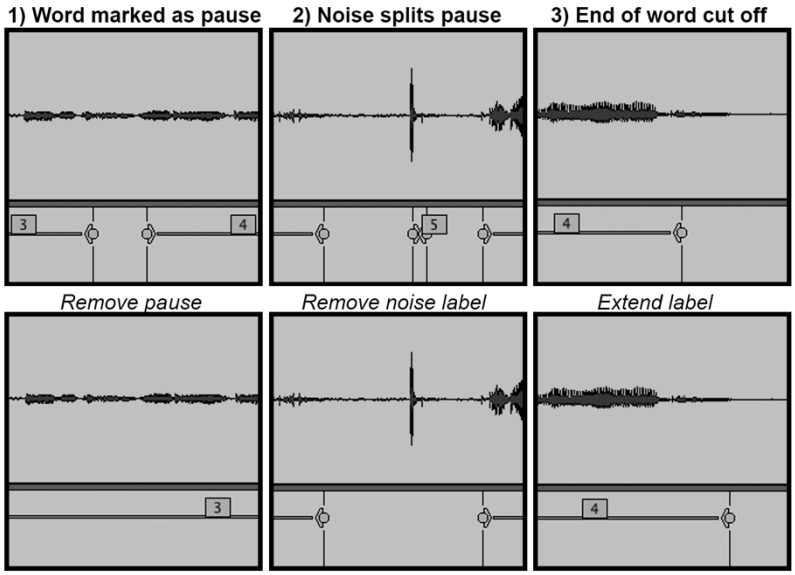

Figure 2. Examples of waveforms that trigger errors in Audacity's automatic pause extraction algorithm, and the actions required to correct the placement of pause labels.

Verifying that all labels were placed correctly required that each track be listened to and labels checked. Once we were satisfied that all labels had been placed correctly, the labeled track was exported as a text file that contained the history of the timepoints where speech bursts began and ended. Pause lengths were then calculated using these timepoints for each participant in each of the five tasks.

Results

Overall, the data analyzed in this study consisted of 3,479 pauses taken over the course of 2.8 hours of recorded speech. Table 1 informs on the quantity of data each participant contributed in each task and on the moments of the associated pause duration distributions. The second column of Table 1 gives the participation rate, and its variation between 63 and 68 reflects the circumstance that speech tasks were added to the study in the first week of its commencement. Columns 3 - 5 give basic participant averaged statistics; people in each task contributed about 10 pauses that together consumed about five seconds in the context of task fulfillment that lasted between 27 and 36 s. Columns 6 - 8 give the first three moments of the pause duration distributions formed by pooling pauses over participants. These distributions are illustrated in Fig. 3.

Table 1

Descriptive statistics for 5 speech tasks

					pause duration distribution moments		
Task:	participants	mean pause count	mean time pausing (s)	mean utterance time (s)	mean (s)	SD (s)	skew
question/answer	63	10.2	5.38	30.0	0.53	0.19	0.86
map	68	10.3	5.60	30.2	0.55	0.20	0.71
picture	68	9.5	5.13	27.4	0.54	0.19	0.89
poem	65	12.5	5.97	28.4	0.48	0.16	1.15
paragraph	64	10.6	5.20	35.7	0.49	0.15	0.92

Insofar as pause duration distributions represent specific examples of timing distributions, the five distributions formed in this study all have positive skew, but the specific distribution shape depends on the whether the task involved composition (question/answer, map, picture) or reading (poem, paragraph). Table 2 shows values of the Kolmogorov-Smirnov D statistic for pairwise distribution comparisons, and it is clear that the two task types form two distinct clusters. Distributions formed from composed speech are distinguished from those formed from read speech, but within a class the distributions are not distinguished. This result is, perhaps, not surprising, simply reflecting the circumstance that people produce longer pauses (500 to 750 ms) when composing speech than when reading.

Table 2

Kolmogorov-Smirnov test statistic D comparing pause distributions from five speech tasks

Task:	question/answer	map	picture	poem	paragraph
question/answer	—	0.065	0.055	0.14***	0.11***
map		—	0.064	0.17***	0.16***
picture			—	0.17***	0.14***
poem				—	0.070†
paragraph					—

†$p < .10$ *** $p < .001$

The focus of this study was on body-size scaling, and these results are presented in Fig. 3 as regressions of mean pause duration against participant height for each speech task. The regressions were computed in the log-log plane, as allometric laws are typically expressed as power laws of mass or body size, and in this plane, the slope is the power law exponent. The regression equation then for each task is

$$\log(\text{mean pause duration}_i) = a + b*\log(\text{height}_i) + e_i,$$

where the subscript *i* denotes the ith participant, *a* is an intercept that is not of present interest, *b* is the slope or power law exponent, and *e* is a residual term. *p* values given in the figure refer to the null hypothesis that allometry is absent and that the true slopes are zero.

The figure makes the statistical case that segmenting pauses satisfy allometries. Across speech tasks, the power law exponents ranged between 0.93 and 1.56, all exponents were significantly different from zero ($p < .001$), and the proportions of variance explained by the regressions ranged between 18% and 36%. In this exploratory study, there was no expectation that exponents or correlations would be meaningfully related to the specific speech task or to whether the task involved read or composed speech, and post-hoc tests did not reveal any significant differences in these measures among the various tasks.

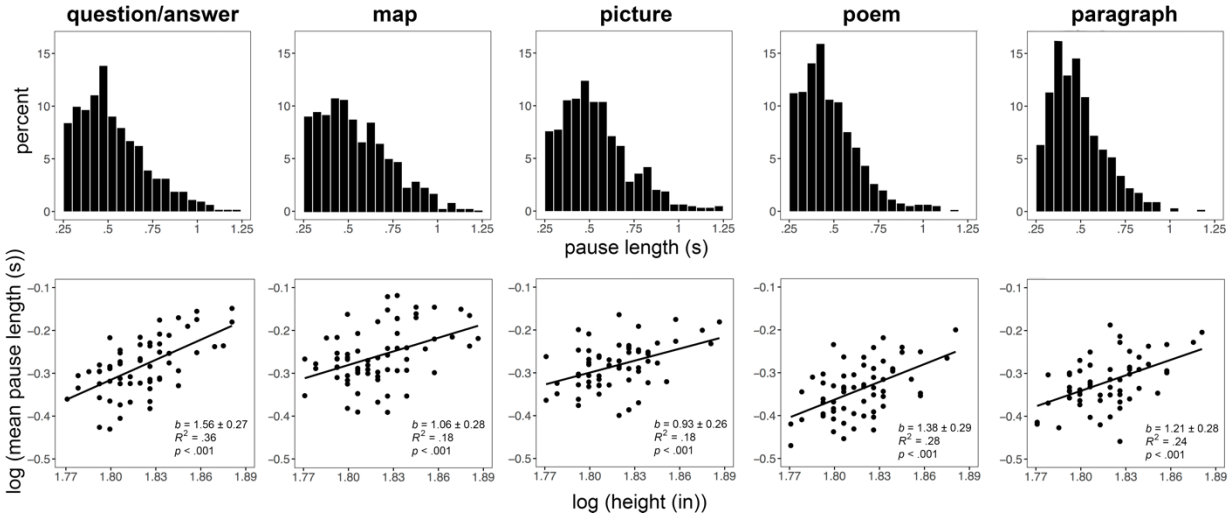

Figure 3. Pause distributions and mean pause length regressions on height for the five speech tasks in the log(inches) – log(seconds) plane.

A concern in the report of a novel finding is that it may have been produced unintentionally through some aspect of the procedure or data analysis. Here there is the issue that the pause labeling algorithm in Audacity was checked for errors by the second author who was aware of the purpose of the study. While height data was kept separate from the audio files, height cues in speech (voice depth) raise the issue of whether label checking could have led to the artificial creation of height trends. We have examined this issue by reanalyzing the question/answer task using pause labels that were not checked for errors. In this instance, the audio was not listened to, and the only corrections that were made were to delete the few aberrant labels that marked speech bursts with zero duration. Results from the two analyses were highly correlated ($b = 0.76 \pm .10$, $p < .001$, $R^2 = .54$), but the unchecked labels analysis produced a slightly weaker allometry with a shallower slope ($b = 1.13 \pm .34$, $p < .001$, $R^2 = .17$). The diminished R^2 when the algorithm is unsupervised is expected, in so far as algorithm errors do inject random uncorrelated noise into the pause record.

In order to get a more global picture of segmenting pauses, we formed pause distributions aggregated over the five speech tasks for each participant. Recognizing that the exponent measured for an aggregate allometry may be considerably influenced by the particular tasks that are aggregated, it was still of interest to determine what an ensemble of tasks might produce collectively. As the basis for this analysis, both means and medians were computed for the aggregate distributions for each participant. The mean aggregated pause allometry, illustrated in Fig. 4, explains 50% of the variance with a power of 1.35±.17. This result is not greatly influenced by the skew in the aggregate pause distributions; the same analysis conducted with medians explained 45% of the variance with a power of 1.28±.18.

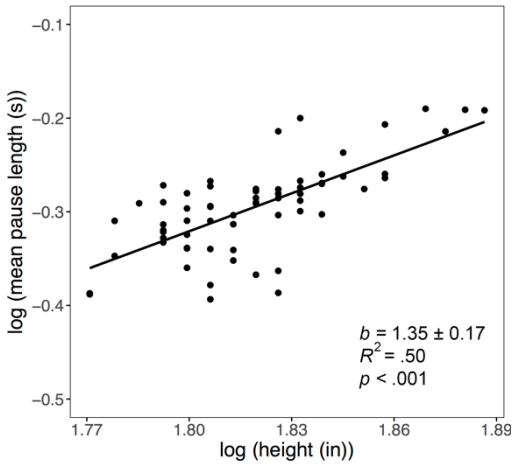

Figure 4. An allometry for mean duration of segmenting pauses formed from the aggregation of pauses across the five speech tasks.

The aggregate regressions may be expressed in a form more common in the allometry literature by rewriting height in terms of fat-free mass (FFM) [Footnote 2]. As FFM varies approximately as height squared in adult humans (Heymsfield et al., 2007), the power law exponents with FFM as a base are respectively 0.68 and 0.64. As these two values bracket 2/3 within the errors, the aggregate allometry may be written simply as

$$(\text{mean segmenting pause}) \sim \text{FFM}^{2/3}.$$

A mass exponent of 2/3 for a duration allometry is objectively large. Some sense of how large 2/3 is can be had by examining the exponents associated with the durations that arise in bodily energy consumption. In human physiology, there are three related durations produced by a body at rest; heart beat period, respiration period, and blood circulation time. All of these acquire scaling from the mass dependency of basal metabolism. Johnstone et al. (2005) developed a multiple regression model of metabolic effort specific to human bodies that predicts that these physiological durations scale as $\text{FFM}^{.38}$. A mass exponent of .38 is some three

standard deviations smaller than the pause duration exponent of .66. There is also a characteristic duration for a body in walking motion that arises from the pendular period of the legs. When a walking motion drives the legs at their resonant frequency, walking is maximally efficient, in that energy consumption is minimized, and the legs find a natural period of oscillation that scales as $height^{1/2}$ or $FFM^{1/4}$. An exponent of ¼ is about 4.5 standard deviations smaller than the pause exponent. To the extent that body energetics provide meaningful examples of physiological and behavioral duration allometries, pause duration allometry clearly exhibits a singular sensitivity to body size.

Discussion

Allometry appears to be a general feature of segmenting pause duration. While the five speech tasks do not represent five independent replications, they did offer different opportunities for speech style, and every speech task produced allometric scaling. Allometry also appears be an important feature of segmenting pause duration in the sense of effect size; the correlation coefficients across speech tasks are all in the range .42 to .60. Furthermore, the aggregate allometry explains 50% of the variance in mean segmenting pause duration, an extraordinary outcome for a single regressor variable in the context of pause phenomenology. In fact, such a large proportion of variance explained suggests that, although we have conceptualized segmenting pauses in terms of linguistic/cognitive structures, the allometry in pause duration is caused by some physical aspect of speech mechanics.

While physical constraints in speech production do set the durations of articulatory pauses, such pauses have durations less than 250 ms, and the rationale for setting a high threshold of 250 ms as the minimum acceptable pause length was specifically to exclude articulatory pauses from the class studied. Within the class of segmenting pauses, the one epoch

where physical constraints might influence pause duration is in the moments when the pause ends with the reinitiation of speech. In those moments, the speaker is actively using the diaphragm, abdominal muscles, chest muscles, and lungs to project an air stream towards the vocal folds.

Given the physical complexity of speech production, it is fortunate that empirical estimates of aspiration onset lags are available through the positive voice onset times (VOTs) for the stop consonants /p/ and /k/. Examination of the VOT literature suggested that VOT is not capable of explaining segmenting pause allometry. First, VOTs for /p/ and /k/ are brief, typically bounded by 75 to 100 ms (Mielke & Neilsen, 2018), and so too small to account for the 200 ms range that height predicts in our aggregate data. Moreover, there is evidence from a developmental study (Yu, De Nil, & Pang, 2015) that VOT is not positively correlated with age. If VOT were positively associated with body size, some hint of this should have appeared in their data. Rather, it appears that VOT is independent of age beyond the age of 12. Both lines of reasoning lead to the common conclusion that the physical act of speech production cannot produce the allometries observed in Experiment 1.

Articulation rate, AR = (total number of syllables)/(total speaking time – total time spent on pauses), is also a factor in speech time consumption, and it is conceivable that allometry in pause duration is aligned with allometry in AR. Articulation is, by its nature, mediated by physical processes and so it potentially offers a second inroad for introducing physical mechanism into pause allometry. Consequently, we have analyzed allometries in articulation rates for the two reading tasks where the total number of syllables was constant across speakers. In distinct contrast with we what found in relation to mean pause length, there was no evidence

of allometry in AR for either reading task. In regression models of AR with height, $R^2 = .03$ ($t(60) = 1.39$, $p < .17$) in poem reading, and $R^2 = .007$ ($t(57) = .65$, $p < .52$) in paragraph reading.

The apparent absence of allometry in AR may reflect important differences between the act of articulation and the act of pausing. In the first place, articulation is a skill, one that eventually becomes highly practiced. Allied with the concept of articulation as skill is the notion that articulation may be disordered, in that it may have characteristics that make it less intelligible or that it deviates from language norms. Pausing, in contrast, is not a skill, it is not practiced in the learning of a language, and it is not an aspect of language production that might be treated by a speech pathologist. If pause behavior is not a skill, then its expression might be expected to be highly idiosyncratic. Similarly, the execution of a highly practiced skill might be expected to find similar expression within a community of people enacting that skill. Evidence for both of these suppositions comes from the coefficient of variation (cv = standard deviation/average) computed over participants. For both reading tasks, cv(AR) = .09. This is arguably an objectively small value, as can be seen when put into the larger related context of Weber fractions where just-noticeable-differences of in the range of 4% to 8% are common. A cv of 9% means that people generally articulate at similar rates. Percentage of time spent pausing showed greater group heterogeneity; cv(PP) = .38 in poem reading and cv(PP) = .26 in paragraph reading. These values are in agreement with those computed by Goldman-Eisler (1968), who assessed the relative contributions of articulation and silence in composed speech arising from interviews. She found that across her 8 participants cv(AR) = .09 and cv(PP) = .33 (we have eliminated one outlier that she included - a person who spent only 4% of their speech time pausing). The conclusion is that individuals express themselves more uniquely with their choice of pause lengths than with their articulation rates. This conclusion has implications for

regression models because homogeneity in a statistic is a form of range restriction. That AR has a cv some three to four times smaller than PP suggests that it is less likely, for purely statistical reasons, to generate a strong correlation with body size.

Articulation behavior and pause behavior also are distinguished fundamentally by the fact they are very different activities. The processes that mediate pause termination are wholly unlike those that determine articulation rate. In the following sections, we develop a cognitive theory of pause duration allometry that is specific to pausing behavior. This theory is based on the human experience of brief intervals of elapsed time.

Timing Theory of Pause Duration Allometry

The finding of allometry in pause duration implies that there is something about the nature of spoken communication, the linguistic aspects of speech that trigger pauses, the cognitive tasks that pauses serve, or the act of pausing itself that somehow couples into the dimensions of the human body. In this section, we construct a theory of how this coupling might be realized. An obvious point of departure in such an open-ended inquiry is the extant literature on the factors that influence pause duration. In this regard, it is known that pause durations are sensitive to a myriad of linguistic factors that include grammatical structure, prosodic structure, discourse organization, phrase length, and phrase complexity (see Ferreira, 1991, 1993; Goldman-Eisler, 1958, 1968; Krivokapić, 2007). The question here is then how such linguistic factors might produce an allometry in pause duration.

One issue that must be reckoned with is that empirical studies of pause duration have tended to be qualitative and exploratory. To illustrate this point we consider the methods employed by Krivokapić (2007). There it is shown, for example, that phrase length impacts

pause duration, but only in the sense of a main effect; phrases divided into the cells *long* and *short* produced the main effect of longer versus shorter pauses. Similarly, it is shown that prosodic complexity impacts pause duration, but again only in the sense of a main effect; intonational phrases divided into cells distinguished by whether they branched into intermediate phrases or not produced the main effect of longer versus shorter pauses. Our appraisal of the empirical literature on pause duration is that the design architecture found in Krivokapić (2007) is generic; studies that manipulate pause duration using experimental designs generally do so through a division of stimulus materials into treatment cells and then to the production of main effects. While this approach is quite common in psychological research, it does not provide the level of measurement precision that would be expected to support a theory of allometry. The basic problem here is that allometries are not framed in terms of main effects. Rather, an allometry is a continuous mapping that makes point predictions, such that for every body size, there is a unique level of some physiological, behavioral, or morphological variable. Although not a proof, it is arguable that an empirics based on qualitative treatment effects does not have the quantitative precision required to explain observations that are based on point predictions.

Beyond issues of measurement is the circumstance that linguistic analyses of pause duration have not been framed in biological terms, being based solely on the formal analysis of speech structure. In order for linguistic theory to be relevant to pause duration allometry, there would have to be an association between body size and the experience of those linguistic factors that influence pause duration. Specifically, this would involve evidence that body size impacts, say, the experience of branching or non-branching intonation phrases. We do not argue that language production and comprehension are disembodied, but it is unclear how linguistic

constructs such as prosodic complexity or phrase length are embodied in the specific way that produces allometry.

A more promising point of departure for developing a theory of pause duration allometry is to consider both the cognitive tasks that pauses serve and the formal requirements for their execution. Krivokapic´ (2007), for example, interprets the main effect that longer pauses are associated with longer phrases as evidence that pauses are used for both the speaker's task of speech planning and for the listener's task of speech comprehension. Consider, then, what is formally required for the successful execution of these tasks. In the planning of subsequent speech, the speaker must select some sort of planning horizon; people do not pause for minutes while they make extensive plans, and this horizon must be negotiated not just in terms of the complexity of planned speech but also in terms of how much time is appropriate for silence while plans are being made. Similarly, speakers must allocate what they judge to be sufficient stopping time to maintain listener engagement and comprehension. Clearly, the successful allocation of time for planning and comprehension requires at a minimum that the speaker be able to discriminate time intervals. Beyond these pragmatic tasks, pauses manifestly also serve the task of giving speech shape, its cadence and rhythms. These rhythms are central to individual expression and involve a finely attuned sense of the duration of silence. It is equally clear then that the production of speech which does not sound mechanical will also require mechanisms that permit the discrimination of time intervals. These considerations suggest that a theory of pause duration allometry might be based upon solutions to the problems of how time intervals are experienced and how they are discriminated. Timing mechanism is not generally a focus in the linguistic analysis of pauses, but here it is potentially relevant as a gateway

psychological process for the introduction of body size scaling. We begin this discussion with a review of how interval timing has been traditionally approached.

Clock Models of Interval Timing

Formal interval timing models are designed to account for the observation that people and other animals are able to produce, reproduce, and estimate targetted intervals of time. In that there is no specific energy associated with time, and no sensory surface for its registration, formal models have mostly been framed in terms of some kind of clocking mechanism. Historically, the most influential of clock models is scalar expectance theory (SET) (Gibbon, 1977; illustrated in Fig. 4 of Meck, Church, & Gibbon, (1985)). In this model, a pacemaker supplies counts such that when they accumulate to a value held in a reference memory, the process terminates with a behavior that signals that a specific interval of time has elapsed. The existence of stored count values in reference memory reflects prior learning and is a critical component of the model. Without stored count values to match, the mechanism cannot terminate count accumulation and so cannot manifest temporal discrimination.

The particular behaviors that SET and other clock models were designed to explain is illustrated by the peak procedure (Roberts & Holder, 1984). In this paradigm, animals (typically rats and pigeons) manifest interval timing by producing maximal response rates (bar pressing or pecking) at an appropriate interval of time passage that has previously been reinforced. This paradigm does invite a model of timing based on matching counts in a reference memory, but it does not describe what people do when they pause speech. Most importantly, rats and pigeons do not choose the time intervals that their behaviors express. Rather, these intervals are chosen by the investigators according to whatever seems reasonable to them. People, however, do choose when to end their pauses and restart speech, and the process is largely improvisational, as

pause ending choices are made in the moment to reflect prevailing linguistic and communication demands. An account of these choices requires a more flexible framework than clock models with reference memories can offer. The theory that is developed below shows how timing behavior can arise without internal clocks, and moreover, how allometry can enter the protocols of time discrimination that underlay pause termination.

A Phase Transition in Human Timing

Our theory of pause termination is motivated by the observation that pause durations are typically brief, less than a couple of seconds. The importance of this observation for any theory of human timing arises from the considerable psychophysical and cultural evidence that the human sense of time operates with one set of processes for short time intervals, less than about 1.5 s, and other sets of processes for longer intervals. In the most elementary terms there is a phase transition in the experience of time. Time intervals less than 1.5 s enter are capable of producing an immanent sensation, a feeling, that supports our awareness of the subjective present. Time intervals greater than about 1.5 s enter awareness more through a process of deliberate reckoning. Fraisse (1984) casts the distinction in terms of time which is *perceived* versus time which is *estimated* with explicit memorial support. To be clear, this language hardly clarifies what is happening cognitively on the short side of the phase transition, as Fraisse's notion of time perceived is not operationally defined. It is exactly the kind of construct that, in the perspective of logical positivism, would be held as not meaning anything. This is not a criticism, however, as there may be no language to describe the human sense of brief, < 1.5 s, intervals of time that does not refer to holistic notions such as "being perceived" or "being felt". There has been some attempt to identify the sense of time on the short side of the phase transition

with interoception (Wittman, 2009), and this may eventually lead to at least an understanding of the physiological correlates of the immanent experience of felt time.

The most singular way in which people experience time passage as a feeling is perhaps through the phenomenon of rhythmic pulse. The emergence of rhythmic pulse is an example of temporal organization where successive beats are brought into relation with one another. An observation which is both globally and historically valid is that beat relatibility is significantly attenuated at tempi below 40 beats per minute (bpm) where the inter-beat interval exceeds 1.5 s. This fact is woven into the fabrication of analog metronomes (e.g. Seth Thomas) where it is simply impossible to position the sliding mass to produce a tempo below 40 bpm. Psychophysical studies of drumming (tapping) performance also confirm that, at tempi slower than 40 bpm, drumming performances tend to meander in tempo (Madison, 2001; Gilden & Marusich, 2009; Gilden & Mezaraups, 2021). The transition that occurs near 40 bpm may be visualized directly in Fig. 5 taken from Gilden and Marusich (2009). This figure depicts the time series of inter-beat intervals (interval between successive drum strikes) for performances in a continuation paradigm (no click track) from a single individual. Each performance consisted of 60 drum strikes after an induction period where the target tempo was introduced. The time series are evidently of two types. At tempi of 40 bpm and faster, the time series are stabilized around the target tempo, while at slower tempi, the time series execute what appears to be a random walk in instantaneous tempo. The interpretation here is that 40 bpm represents a critical point separating a phase of rhythmic feel from a phase of perceiving nothing more than a succession of unrelatable beats. This phase transition and how it is measured is discussed at length in Gilden and Mezaraups (2021).

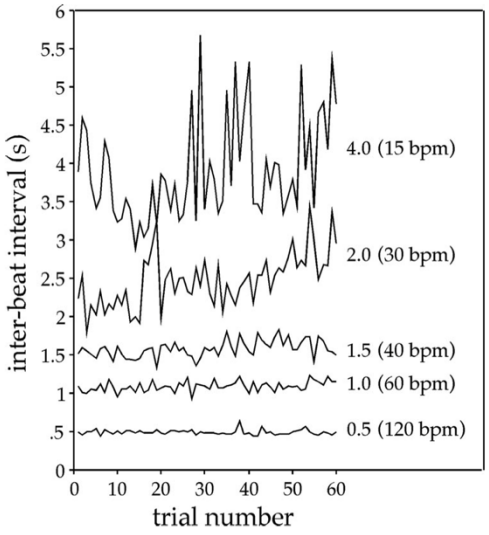

Figure 5. Time series of inter-beat intervals for drumming performances at five tempi.

Psychophysical studies of temporal interval discrimination conducted by Grondin, Meilleur-Wells, and Lachance (1999) have identified essentially the same phase transition. They measured the critical interval at which explicit counting improves discrimination performance in a two-interval forced-choice paradigm (2IFC). Explicit counting lowered difference thresholds at intervals of 1.18 s and larger. This result implies that intervals less than about 1.2 s can be sensed without any kind of external or explicit support, and that whatever is the nature of that sensing, it can be held in memory for the purposes of comparison. That the 2IFC transition occurs at slightly lower value than the pulse transition may be due to the working memory demand that exists in 2IFC methods generally. Nevertheless, the two estimates of the transition point are in good agreement, and it deserves to be noted that musicians will often subdivide the interval between beats at slow ballad tempo – 50 bpm (inter-beat interval of 1.2 s).

The observation that there is a phase transition in how time is experienced would seem to be sufficiently salient and noteworthy that it would play a large role in theoretical models of

timing. Yet, for the most part, timing models do not recognize the existence of two distinguishable phases. SET provides a good example of a clock model that operates on the presumption that there are no real distinctions that exist in the experience of time intervals other than, perhaps, some intervals are longer or shorter than others. In SET, the accumulator does not have any internal structure that creates a distinction in count number beyond numerosity. Similarly, there is no distinction in the values held in reference memory beyond value magnitude. So while clock models of interval timing do have the virtue of being computationally explicit, they also suffer the cost of being unable to recognize the most fundamental distinction in human temporality. The theory outlined below, in. contrast, takes as a point of departure the existence of a phase transition and the empirical fact that speech pauses are typically found in just one phase.

Fullness of Time Theory of Timing

As pause durations are generally located on the short side of the phase transition, in the regime of temporal experience where it makes sense to refer to felt or perceived time, the central construct in our theory of pause termination is the feeling of pause fullness. While this construct is not itself directly observable, it does lead to an analytic theory of pause termination with empirical entailments. Fig. 6 illustrates the theory through a depiction of the anatomy of a pause. The pause begins with the reception of a signal that causes speech to halt. This signal might be a grammatical or prosodic boundary, it might be an interruption, it might be losing one's train of thought. Whatever causes speech to halt, increases in pause duration in real time are coordinate with an increasing feeling of pause fullness, denoted as f(t). The pause is terminated when a state of pause fullness arrives that leads to the choice to recommence speaking. In this way, a mechanism capable only of monitoring states of feeling is able to behave as if it is measuring

elapsed time. Specifically, in Fig. 6, a pause is initiated at t_0 with a fullness level of 0, and is terminated when the sense of fullness builds to the level f_1. The process produces an elapsed time, a pause duration, of $\Delta = t_1 - t_0$ without requiring any explicit representation of time per se. The pause duration Δ is produced only as a by-product of the decision to terminate the pause when the fullness level reaches f_1.

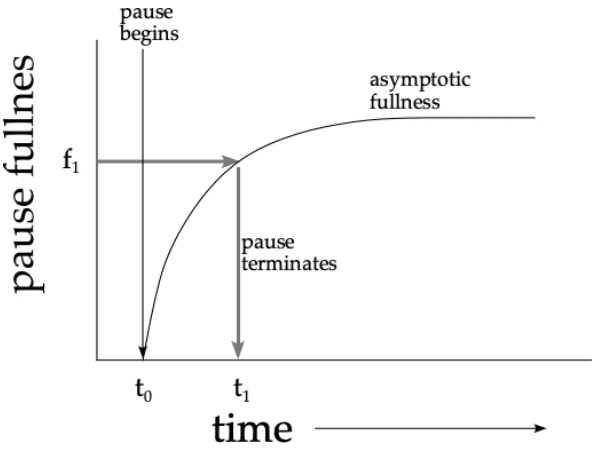

Figure 6. Anatomy of a pause. Pause onset at t_0 initializes an epoch of sensing the fullness of elapsed time. The pause ends and speech recommences when the process arrives at the state of fullness f_1.

The pause fullness function has purposely been conceptualized to asymptotically converge to a maximum level of the feeling of fullness. There are two reasons for this. First, the functions that describe the mapping between stimulus magnitude (intensity) and perceived magnitude generally have upper bounds that are realized as asymptotes. Secondly, asymptotic convergence has the formal property that it creates an upper limit to the intervals of time where pause fullness is useful for temporal discrimination. Such a maximum duration is required by the theory in that the construct of pause fullness is only defined in the regime of felt time, on pauses with durations less than a couple of seconds.

The anatomy of a pause given in Fig. 6 describes a framework for how people experience the passage of brief intervals of time. As a theory of speech pauses it is incomplete in that it contains no reference to linguistic factors such as prosodic complexity or phrase length that are known to influence pause duration. These terms implicitly enter the theory by setting the fullness states that determine pause termination. Pauses taken prior to longer speech phrases, for example, will be associated with larger values of f(t) than pauses taken prior to shorter phrases. As f(t) is conceptualized to be monotonic increasing, larger values of f(t) are associated with longer pauses in real time. In this conception of pause termination, linguistic context is subordinate to the processes that underlay time discrimination. There is an alternative view and that is that linguistic factors introduce their own protocols, and pauses are terminated when these protocols are completed regardless of how the speaker senses time passage. In order to amplify these two views we consider how pause durations may be distinguished from reaction time latencies, the elapsed time between stimulus presentation and response that forms the principal dependent variable that drives cognitive theory.

In typical methodologies, reaction time latencies provide insight into mental structure and function through the assumption that the latency reflects the sum of completion times for a set of well-defined processes. For example, in a visual search methodology, a reaction time latency for a display consisting of a single element would be conceptualized as the sum of completion times for the element to be perceived (about 150 ms), for a decision to be made about whether the element is a target (about 50 ms), for response mapping if a keyboard is being used to collect responses (about 100 ms), and finally for depressing the appropriate key (about 100 ms). Nowhere in this conception of a reaction time latency is there any mention of how the participant experiences time. Here the participant is conceived to be a mechanism that executes visual

search through the rote execution of a specific set of processes. Mechanisms do not need to experience time in order to produce time dependent behavior, and there is no need in theories of mental process based on reaction time measurement to refer to felt time. That the participant is conceived to be a mechanism is what allows visual search behavior, in particular, to be modeled using Monte-Carlo methods of simulation (see for example Thornton & Gilden, 2007).

The need to take time for speech planning is often cited as a global factor that determines pause duration and the question arises whether speech planning times might be viewed as a form of reaction time latency. Consider then what is happening when a person pauses, say at a prosodic boundary where speech planning might be expected to occur. If the pause is in fact used for speech planning does that mean that a set of processing routines are initiated at the beginning of the pause such that when they are completed the pause ends and the planned speech is executed? If this is the case then speech planning times are process completion times, essentially reaction time latencies, and there is no need to refer to how time is experienced or to pause fullness states. This view, however, is based on assumptions that are unlikely to be met in natural speech.

In order for reaction time measurement to produce meaningful data it essential that the participant execute the assigned task and only the assigned task. It is this requirement that allows the latency to be interpreted in terms of a fixed sequence of processing stages. Practically, in order to get a person to behave as a mechanism executing such a sequence, reaction time measurement is conducted within the methodology of speeded forced choice. That it is speeded is critical. If the participant is allowed to terminate the trial in a relaxed and unhurried manner, then the reaction time latency is exposed to factors that are outside of the intended process sequence. These factors are diverse and may include everything from double

checking to ensure high accuracy to producing the inner narrative that forms the stream of consciousness. It is inevitable that a reaction time methodology that is not speeded allows the participant to introduce their sense of time passage, their temporality, into the machinery of perception/categorization/response-mapping/execution. Manifestly, speech production is not speeded in the sense of psychophysical methodology, and so pause durations cannot be construed as a sum of completion times that is independent of the speaker's general sense of time passage. Rather, we have argued throughout that pauses in speech are instrumental in giving speech its rhythms and cadence so that even when a pause allows time for speech planning, it nevertheless is embedded in a communication context that requires temporal discrimination. Consequently in the theory presented here, the processes of temporal discrimination determine the architecture of pause completion while linguistic factors act within the architecture by setting parameters values, values of f(t).

Fullness of Time Function

The path way to pause allometry is through the fullness function that maps clock time into the feeling of the fullness of time. In this section, we construct a fullness function f(t) that has the properties illustrated in Fig. 6, as well as the property of body size scaling in the termination of pauses. A standard model in perceptual decision making, one that is particularly well suited to describing the time course of the feeling of time passage (Toso et al. 2021), is the leaky integrator:

$$df/dt = -f/\tau + s$$

where f(t) is the instantaneous feeling of pause fullness, s is the rate at which the feeling of fullness is supplied, and τ is a leakage or, as will be referred to here, decay time scale. We will make one simplifying assumption in order to solve this equation, and then another to illustrate a

special case where pause termination obeys allometry even when the community of speakers is experiencing an invariant sense of pause behavior. Assuming that the fullness source, s, is constant, this equation may be integrated to yield

$$f(t) = f_a(1 - e^{-t/\tau}),$$

where $f_a = s\tau$. This solution has the behavior of the fullness function illustrated in Fig. 6: The feeling of fullness grows from zero at t = 0, when the pause begins, to an asymptotic level of fullness, f_a, over several decay lifetimes, τ.

The manner in which allometry enters into pause termination times will be clarified by inverting the fullness function so that clock time is a function of pause fullness:

$$t(f) = \tau \ln(f_a/(f_a-f)).$$

This equation states that pauses of duration t(f) will be produced when pauses are terminated at feelings of fullness f. The segmenting pause allometries reported in the first study are formally properties of t(f), and the theoretical issue is what aspect of t(f) lends itself to allometry. There are three constructs in t(f): the decay time scale τ, the feeling of pause fullness, f, and the fullness supply rate, s, that enters through $f_a = \tau s$. While f and s might exhibit allometry, there is nothing in the development of the theory that requires that they have parametric dependencies. The decay time scale, τ, on the other hand, must have parametric dependencies, and so is the natural candidate in this theory for carrying allometry into t(f). Although time scales and scales in general are frequently encountered in physics, they are not often encountered in psychological theory, and some background may be necessary to understand the significance of τ.

The theoretical significance of τ arises from its appearance in f(t) as a time scale in the negative exponential. Time scales play a dual role in physical theories; they are in every instance observed as a specific interval of time, but more importantly, they are also mappings

that carry system parameters into time intervals. An illustrative example is Newton's law of cooling, where substances cool according to a negative exponential. Here, the relevant system parameters are heat capacitance and surface area, and for coffee, the cooling time scale is measured in minutes. Similarly, radioactive decay is formally a type of cooling, in that it is also described by a negative exponential, but with system parameters defined in terms of quantum transition probabilities, the decay time scales may run to thousands of years. In general, where there is an appearance of a scale in an exponential, there is an associated piece of physics that specifies a function of system parameters, $\tau(a_1,a_2,a_3,\ldots)$. It is this function, together with the specific values of its arguments that determines the rate of real time system evolution. In the present context, if body size is a parameter of τ, and if τ(body size) is a monotonic increasing function of body size, then pause durations will show the allometries discovered in Experiment 1.

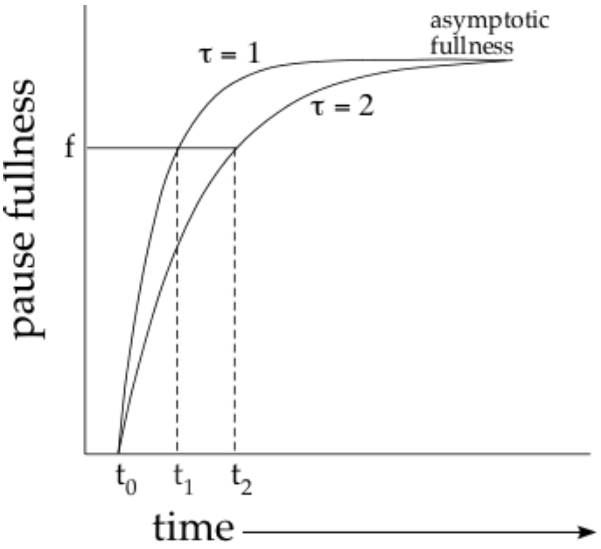

Figure 7. Illustration of how allometry in pause termination may arise even when the psychological experience of time passage is universal.

Fig. 7 graphically illustrates the mathematical theory of how allometry in pause duration is created by body size dependence in τ. Depicted is a limiting case where pause durations satisfy an allometry, even when the psychological experience of speech is invariant over speakers. Pause fullness growth functions, $f(t) = f_a (1 - e^{-t/\tau})$, are shown for two speakers, one (shorter) with a decay time scale $\tau = 1$, and the other (taller) with $\tau = 2$. In this example, both speakers share the same asymptotic state, f_a, and so experience the same range of pause fullness states. For the purpose of illustration, consider that that these two speakers encounter an opportunity for a pause at t_0. Each speaker pauses until the same sense of fullness, f, arrives before they proceed with more speech. In real time, however, the taller speaker with the longer decay time scale terminates their pause at t_2, while the shorter speaker with the shorter decay time scale terminates their pause earlier at t_1. The inequality relation

$$t_2 > t_1 \text{ if speaker-size}_2 > \text{speaker-size}_1$$

is the mathematical restatement of the allometries found in Experiment 1.

Experiment 2: Allometry in Long Pauses

The theory that leads to t(f) allometry has generality beyond the pause behavior associated with segmenting pauses. It should apply to any pause behavior where durations are less than a couple of seconds, the domain where the theory of pause fullness is defined. In order to extend the empirical support for the theory beyond segmenting pauses, we consider a class of pauses that are universally encountered but little discussed in the pause literature. The pauses of interest here are those terminated by a filled pause, typically “um” or “uh”, and acquire class definition from the linguistic functions of filled pauses. Filled pauses carry a variety of

meanings, one of which is acknowledgement that a delay is in progress, and that speech will soon commence (Clark & Fox Tree, 2002). The meaning of a filled pause is quite important here, because acknowledging that a delay is in progress is a judgement about how the delay feels, that it feels long. It is not an acknowledgement that arises from counting to 10 or from a glance at a clock. In this way, filled pauses that terminate long planning pauses keep the duration of planning pauses in the regime of felt time. As the theory of allometry is intended to apply to assessments of felt time generally, it must also extend to long planning pauses that are terminated by filled pauses.

In the following study, further distinction from segmenting pauses was achieved by considering only filled-pause-terminated-pauses that occurred between the asking of a question and its answering. Such pauses have the property that they occur outside of speech, being initiated by a question and then being terminated prior to the delivery of an answer. This refinement guarantees that our new class of pauses is not associated with any of the linguistic landmarks that create segmenting pauses. As a consequence, this class of planning pauses presents a meaningful opportunity for not finding allometry and so for falsifying the theory.

In addition to the prediction that this new class of pauses will also obey allometry, the mathematical development of the theory permits a stronger, second, prediction. The expression for pause length is the product of two separate components: $t(f)= \tau \ln(f_a/(f_a-f))$. In a given speaker, τ is regarded as being fixed, and pauses of different lengths are created by different feelings of time fullness, f, or fullness contrast f/f_a. If allometry in t(f) enters through allometry in τ, then a way of thinking about this expression is that

t(f) =(body size sensitive piece) x (pause length setting piece).

This expression suggests that allometric properties should be independent of whether the pauses are segmenting and relatively short, or terminated by a filled pause and relatively long. As allometries only have one parameter, the body size exponent, the theory makes the prediction that the new class of pauses will obey an allometric law with an exponent representative of the exponents derived for segmenting pauses.

Method

Participants

Forty-four native English speakers were recruited from the undergraduate subject pool at the University of Texas at Austin and received course credit for their participation. Ages ranged from 18 to 25 years, and heights were between 62 and 75 inches, with a median of 68 inches.

Stimuli

The study consisted of responses to the following 11 questions that were chosen to invite a substantial pause prior to the beginning of response: What is your favorite thing to do over the holidays? What is your relationship like with your family? What do you look for in a presidential candidate? What would your ideal date be? What is your favorite thing to cook? What is your dream career? Do you prefer truth or beauty? Defend your answer. Do you think that robots can make art? How do you feel about e-scooters on campus? Who was your childhood hero, and why? What do you think happens when you die? All questions and answers were recorded continuously using the same equipment described in Experiment 1.

Procedure

Participants were seated in a quiet room and situated directly in front of the microphone and computer. The experimenter first explained that they would be asking a series of questions and that audio would be recorded continuously. It was emphasized that they should respond

naturally and that there were no right or wrong answers. Once it was clear that the participant understood the task, the researcher started the recording in Audacity and read each question aloud one at a time. Questions were always asked in the same order, and the participant was given ample time to answer each one before moving on to the next. The researcher did not respond or comment to any of the answers and attempted to maintain a neutral expression.

Pause Extraction

The Label Tool in Audacity was used to mark the time points at which pauses began and ended. Two types of pause were marked and exported for data analysis: *pause before fill* that occurred following a question and prior to answer commencement, and *filled* pauses (such as "um" and "uh") that interrupted such initial silent pauses. To be clear, both types of pause were marked only when they occurred consecutively and directly after a question.

Results and Discussion

In this study there were 11 questions and consequently there were 11 opportunities for the introduction of a planning pause following a question that was terminated by a filled pause prior to the commencement of an answer. On average each participant contributed 3.4 pauses of both types so that the entire data set consisted of about 150 pauses of both types The pause distributions and height regressions from this study are illustrated in Fig. 8 and the distribution moments for the two pause types are listed in Table 3. We will consider the distribution results first.

Table 3

Pause duration distribution moments

Type:	mean (s)	SD (s)	skew
pause before fill	0.96	0.37	0.37
filled pause	0.58	0.13	1.02

Referring to Fig. 8, the most salient aspect of the pause before fill distribution is that there are few of these pauses that exceed 1.5 s in duration. This is exactly what should be observed if speakers were terminating these pauses with a filled pause so as to keep them on the short side of the phase transition where pause fullness and the feeling of time are experienced. Also noteworthy is that the experimental design succeeded in displacing this pause distribution from the segmenting pause distributions shown in Fig. 3. It is centered at 1 s, where segmenting pauses are rare, and it depopulates below .5 s which is the mean of the segmenting pause distributions.

The distribution of filled pauses durations is of interest in that it informs on the extent to which words such as “um” and “uh” were prolonged in the question/answer context created in this study. In the distribution shown in Fig. 8 over half of the filled pauses have durations larger than .5 s and so should be viewed as prolonged. This classification is important in view of the claim that prolongation of a filled pause indicates that the speaker recognizes that preceding pause was getting a little long (the prolongation hypothesis of Clark & Fox Tree, 2002). So long, in fact, that something needed to be said about it, hence the utterance of an “um” or “uh”. Also noteworthy is that the filled pause duration distribution rapidly depopulates beyond .75 s so that the distribution is effectively truncated at 1 s. That all of the filled pauses are held on the short side of the phase transition suggests that that prolongations have the property of growing fullness in common with silent pauses.

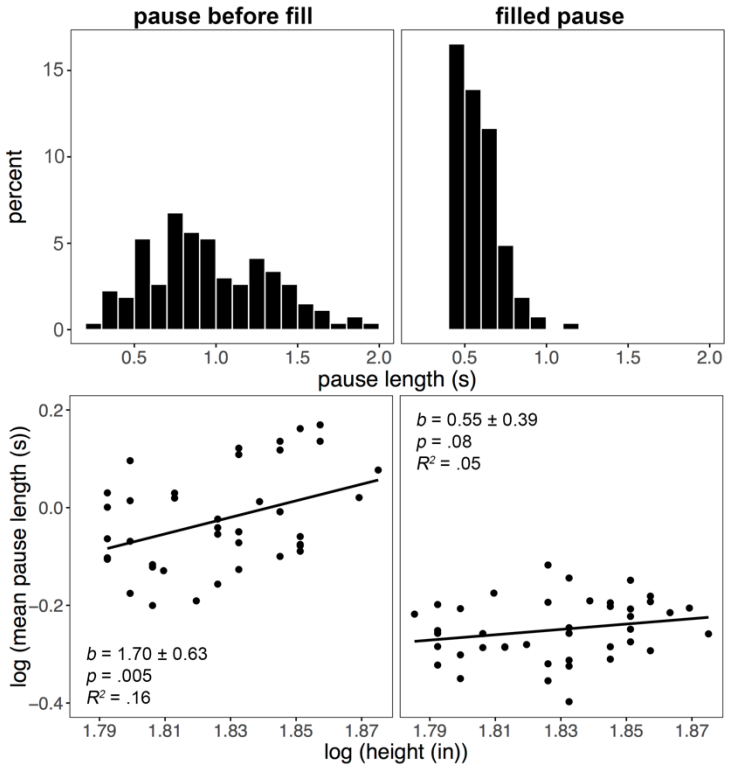

Figure 8. Pause distributions and mean pause length regressions on height. The first column shows results for pauses taken prior to the answering of a question that were terminated by a filled pause. The second column shows results for the associated terminating filled pauses.

The regressions shown in Fig. 8 provides evidence for allometry in long planning pauses and marginal evidence for allometry in the prolongation of filled pause words. The theory predicted that long planning pauses would obey allometry to the extent that their durations were held on the short side of the phase transition, and this is observed. Beyond the theory surviving a critical test, the empirical production of a second instance of allometry in pause duration in the regime of felt time is empirically important. The demonstration of allometry in pause duration is obviously novel, and it is the case that novel findings, while interesting, have what Ioannidis (2005) refers to as “low pre-study odds”. Novel reports from a single experiment must create

concern that the findings will not replicate. A second independent instance of an allometry in pause duration lends considerable rhetorical force that the empirics are sound. The theory also made a point prediction, that the planning pause exponent would be representative of exponents measured for segmenting pauses. This prediction was also verified; the derived exponent for long planning pauses, $b = 1.70 \pm .63$, is in the same range as measured for segmenting pauses. This result is important in view of the patent fact that pause durations are not produced by a physiological process, and there is no theoretical basis for predicting the values of these exponents. Yet, the theory is able to predict that exponents in different regimes of pause duration will be similar, and in the case investigated, this also is observed.

The weak evidence for allometry in filled pause duration requires some discussion. Intuitively, the decision to terminate a prolonged filled pause does not appear to be materially different than the decision to terminate a silent pause of equivalent duration. The theoretical development that led to the prediction of allometry in t(f) could equally be applied to prolonged filled pauses. That this expectation has not been realized may be due to both restriction of range and also to the presence of non-prolonged filled pauses. With regards to restriction of range, the relevant issue is how much of the available range of felt time is populated. Over the two studies reported here, pause durations were observed to occupy a range that extended from .25 s (the boundary with articulatory pauses) to about 1.5 s (the boundary where felt time ends and estimated time begins). Segmenting pauses populate between .25 s and about 1 s, with coefficients of variation (standard deviation/mean, calculated from Table 1) of about .35 with little dependence on task. Planning pauses that are terminated by filled pauses populate between .3 s and 1.5 s, with a cv = .39 (calculated from Table 3). Filled pauses populate the narrowest band, from about .4 s to .9 s, and with a cv = .22 (calculated from Table 3). Here we do not

regard restriction of range in filled pause duration as a statistical artifact that might be remedied or corrected, but rather as simply an observation that participants in our study did not choose to prolong their “ums” and “uhs” out to 1 s and beyond. Whatever the reason for this restriction, it has the statistical entailment that regression effects generally would be expected to be suppressed for filled pauses relative to regression effects for segmenting and planning pauses.

A second issue is that a number of the filled pauses had durations less than .5 s, and many of these might not be judged to be prolonged. The distinction between filled pauses that are spoken at normal rates of articulation and filled pauses that are deliberately prolonged is important here. Only the latter are presumably terminated by choices based on how the time spent in prolongation feels. As the theory of pause fullness only applies when pauses are terminated on the basis of felt time, when a filled pause has its duration set by articulatory mechanisms, the feeling of the fullness of time ceases to be relevant, and allometry does not arise as a prediction. To the extent that our sample consists of a mixture of both prolonged and non-prolonged filled pauses, any regression effects that exist only for the prolonged component will be diluted.

General Discussion

This article provides evidence that the pauses taken in the course of fluid speech and in a class of long pauses terminated by a filled pause obey allometry. Allometry is well-known in biology for describing aspects of physiology and morphology, but here it describes an aspect of how time is negotiated in communication. Body size scaling of pause duration is not a small effect and was evident in each of the five speech tasks assessed in the first study. The proportion of variance explained by body snize in the aggregated pause corpus was .50, a value that is quite

large for a single regressor in linguistics or psychology more generally. These findings motivated a theory of pause duration that gave the body a role to play.

The theory of pause duration presented in this article is in no way constructed from first principles, but it does provide an account for the finding that body size does influence pause behavior, a finding that has no precedent and therefore no prior theoretical treatment. As the theory departs significantly from clock models of timing, we summarize here its key assumptions and points of development. The theory begins with the recognition that pause durations reflect the activity of a memory process that is sensitive to time passage. This recognition provided little theoretical direction, however, in view of the circumstance that the sensing of time is an open problem in psychology and neuroscience, notwithstanding the large literature on clock models. A second observation of note is that decisions to terminate pauses are made in the moment, and a theory of how this is done must be capable of explaining how temporal discrimination can operate in a fluid improvisational environment. Our theory of time discrimination approached the in-the-moment nature of pause termination by invoking a continuous mapping between distal time and the feeling of time passage. This mapping allows pause ending decisions to be made in the moment without recourse to the architectural assumptions of scalar expectancy theory – pacemakers, accumulators, and reference memories. Such a solution, however, only makes sense if people do, in fact, feel the passage of time. While the timing literature is certainly not about how time feels, that fact should not call into question the proposition that people universally do feel time, and that one of the most common vehicles for this experience is rhythmic pulse. The fact that people lose the sense of pulse when beats are separated by more than 1.5 s suggests that this value marks the boundary between felt/perceived time and explicitly estimated time. Further development of the theory to explain why pause

durations obey allometric laws required setting out a specific mathematical framework for the growth of feelings of pause fullness. This demand was met by a leaky integrator model of fullness growth that had a negative exponential appearing in the solution. The decay lifetime, τ, that scales time in the exponential is the core element of the entire theory. It supplies through its parametric dependencies the connection between world time and the interior sense of time passage. The conjecture that a system parameter of τ is body size completed the theory.

The theory is quite abstract insofar as it makes reference to constructs such as decay time scales and pause fullness supply rates that cannot be directly measured, but it does make two predictions. The predictions follow from the circumstance that, while the theory was developed to explain pause behavior, it contains no terms that relate specifically to language or communication. Consequently, the theory predicts that allometry should be observed generally in behavior that is based on the discrimination or production of felt time intervals. The mathematical expressions that were derived in solving the leaky integrator differential equation led to a second prediction, that allometric exponents would be independent of the regime of felt time that the behavior was expressing.

A second experiment investigated a type of pausing behavior that was intended to be independent of the production of segmenting pauses that occur at prosodic and grammatical boundaries in fluid speech. People also pause when planning responses to difficult questions, and when such pauses approach a point where it is clear that a delay is in progress, they will often emit a filled pause – an “um” or “uh”. In this context, the pauses that are produced during speech planning and which are terminated by a filled pause were also found to satisfy an allometry, and with an exponent that was in the range of derived exponents for segmenting pauses. To this extent, the second experiment extended the range of pausing behavior where

allometry is observed, and it suggests that the derived exponents might be characteristic of behaviors that involve an awareness of time passage.

Finally, there are several respects in which the allometries derived here for pause durations are distinguished from those derived in biology. Foremost is that allometries in the latter fields generally involve a relationship between physical quantities like heart period, head size, or burst acceleration with the physical quantity of body size. In many cases, the physical grounding of all the terms that appear in the allometry permits the construction of a geometric theory that relates body size to whatever biological variable is under consideration. The geometric relation is referred to as an *isometry*, while the term *allometry* is reserved for any relation that is not reducible to geometry. In such contexts, any empirically derived exponent may be compared with whatever a theory based just on geometry would predict. The isometry effectively acts as a kind of null hypothesis. Taking the Kleiber Law as an example, homeostatic temperature regulation requires that the total resting metabolic rate balance radiative losses at the body surface. This statement in itself is sufficient to develop an isometric theory of the scaling of metabolic quantities based on body surface area. The isometric law for basal metabolism is

$$\text{metabolic rate (watts)} \sim \text{mass}^{2/3},$$

and this provides contrast and context for the Kleiber Law (Kleiber, 1932),

$$\text{metabolic rate (watts)} \sim \text{mass}^{3/4}.$$

The isometric exponent has been invaluable both in theory development of how biological systems are organized (West & Brown, (2005), as well as in guiding refinements in the regression analysis through which exponents are derived (Heusner, 1982).

The interpretation of pause allometry exponents, in contrast, is not supported by an isometric relation. Were pause durations set by articulatory constraints, it is conceivable that the

physical properties of the articulatory apparatus would define an isometry. However, pause durations longer than about ¼ s reflect myriad cognitive processes including those that underlay the discrimination of bits of felt time, and as there is no physical principle that connects felt time to body geometry, it does not appear that there will be an isometric theory of pause duration.

In the absence of isometric exponent, the exponent magnitudes that have been produced here through regression analysis can only be evaluated in terms of the standard error of the estimate. This limitation permits the usual statistical questions – is the exponent different from zero, or are exponents derived from different conditions or different studies distinguishable. But unlike the ¾ exponent of the Kleiber Law, it is not possible to say that a particular exponent is interesting and demands a deeper theoretical perspective than geometry can provide. This situation is typical of cognitive psychology, where the magnitudes of measured quantities are evaluated empirically in terms of the variation within an experimental design or in terms of the variation across studies. A deeper understanding of the derived exponents will involve expanding the speech tasks, generalizing to a wider range of speakers, and finally generalizing to other languages. The results reported here should be viewed as the first efforts to understand a new aspect of language behavior.

References

Bransford, J. D., & Johnson, M. K. ([1]972). *Journal of verbal learning and verbal behavior*, *11*(6), 717-726.

Butcher, A. (1981). Aspects of the speech pause: Phonetic correlates and communication functions. *Arbeitsberichte Kiel*, (15), 1-233.

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. In *Speech prosody 2002, international conference*.

Dalton, P., & Hardcastle, W. J. (1977). Cluttering and disfluency of organic origin. *P. Dalton & WJ Hardcastle Disorders of fluency and their effects on communication*, 107-124.

Demol, M., Verhelst, W., & Verhoeve, P. (2007). The duration of speech pauses in a multilingual environment. In *INTERSPEECH-2007*, 990-993.

Feldhütter, I., Schleidt, M., & Eibl-Eibesfeldt, I. (1990). Moving in the beat of seconds: analysis of the time structure of human action. *Ethology and Sociobiology*, *11*(6), 511-520.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*, 210–233.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review, 100(2)*, 233–253.

Gerstner, G.E., & Cianfarani, T. (1998). Temporal dynamics of human masticatory sequences. *Physiology & Behavior, 64*, 457-461.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84*, 279-325.

Gilden, D. L., & Marusich, L. R. (2009). Contraction of time in attention-deficit hyperactivity disorder. *Neuropsychology*, *23*(2), 265.

Gilden, D. L., & Mezaraups, T. M. (2021). Allometric scaling laws for temporal proximity in perceptual organization. In press at *Psychological Review.*

Goldman-Eisler, F. (1958). The predictability of words in context and the length of pauses in speech. *Language and Speech*, *1*(3), 226-231.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. New York: Academic Press.

Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1966). Sequential temporal patterns in spontaneous speech. *Language and Speech*, *9*(4), 207-216.

Heusner, A. A. (1982). Energy metabolism and body size I. Is the 0.75 mass exponent of Kleiber's equation a statistical artifact?. *Respiration physiology*, *48*(1), 1-12.

Heymsfield, S. B., Gallagher, D., Mayer, L., Beetsch, J., & Pietrobelli, A. (2007). Scaling of human body composition to stature: new insights into body mass index–. *The American journal of clinical nutrition*, *86*(1), 82-91.

Hieke, A. E., Kowal, S., & O'Connell, D. C. (1983). The trouble with" articulatory" pauses. *Language and speech*, *26*(3), 203-214.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, *2*(8), e124.

Johnstone, A. M., Murison, S. D., Duncan, J. S., Rance, K. A., & Speakman, J. R. (2005). Factors influencing variation in basal metabolic rate include fat-free mass, fat mass, age, and circulating thyroxine but not sex, circulating leptin, or triiodothyronine. *The American Journal of Clinical Nutrition*, *82*(5), 941-948.

Kien, J., & Kemp, A. (1994). Is Speech Temporally Segmented? Comparison with Temporal Segmentation in Behavior. *Brain and Language, 46*, 662-682.

Kleiber, M. (1932). Body size and metabolism. *Hilgardia*. *6 (11)*, 315–351.

Krivokapic´, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics, 35,* 162-`179.

Madison, G. (2001). Variability in isochronous tapping: higher order dependencies as a function of intertap interval. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(2), 411.

Meck, W. H., Church, R. M., & Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology: Animal Behavior Processes, 11*, 591-597.

Mielke, J., & Nielsen, K. (2018). Voice Onset Time in English voiceless stops is affected by following postvocalic liquids and voiceless onsets. *The Journal of the Acoustical Society of America*, *144*(4), 2166-2177.

Po, J. M. C., Kieser, J. A., Gallo, L. M., Tésenyi, A. J., Herbison, P., & Farella, M. (2011). Time-Frequency Analysis of Chewing Activity in the Natural Environment. Journal of Dental Research, 90(10), 1206–1210. https://doi.org/10.1177/0022034511416669

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in cognitive sciences*, *1*(2), 56-61.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of experimental psychology: Learning, Memory, and Cognition*, *21*(4), 803.

Roberts, S., & Holder, M. D. (1984). What starts an internal clock? *Journal of Experimental Psychology: Animal Behavior Processes, 10,* 273-296.

Schleidt, M. (1988). A universal time constant operating in human short-term behaviour repetitions. *Ethology*, *77*(1), 67-75.

Schleidt, M., & Feldhütter, I. (1989). Universal time constant in human short-term behavior. *Naturwissenschaften*, *76*(3), 127-128.

Schleidt, M., Eibl-Eibesfeldt, I., & Pöppel, E. (1987). A universal constant in temporal segmentation of human short-term behavior. *Naturwissenschaften*, *74*(6), 289-290.

Suess, Dr. (1960). *One Fish, Two Fish, Red Fish, Blue Fish*. New York: Beginner Books.

Simon, H. A. (1966). A note on Jost's law and exponential forgetting. *Psychometrika*, *31*(4), 505-506.

Thornton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review, 114*, 71–103.

Toso A, Fassihi A, Paz L, Pulecchi F, Diamond ME (2021) A sensory integration account for time perception. *PLoS Computational Biology, 17(1)*:e1008668. https://doi.org/10.1371/journal. pcbi.1008668.

West, G. B., & Brown, J. H. (2005). The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. *Journal of experimental biology*, *208*(9), 1575-1592.

White, P. A. (2017). The three-second “subjective present”: A critical review and a new proposal. *Psychological bulletin*, *143*(7), 735.

Yang, L. C. (2004). Duration and pauses as cues to discourse boundaries in speech. In *Speech Prosody 2004, International Conference*.

Yu, V. Y., De Nil, L. F., & Pang, E. W. (2015). Effects of age, sex and syllable number on voice onset time: evidence from children's voiceless aspirated stops. *Language and speech*, *58*(2), 152-167.

Footnotes

[1]An example drawn from sentence construction may clarify the interplay between proximity constraints and grouping. It is common in studies of memory that participants are exposed to lists of words, and it is tacit in these paradigms that the lists be heard as lists and not as strange sentences. In Roediger and McDermott (1995), for example, participants were exposed to the semantic associates of "spider"; "web", "insect", "bug", "fright", and so on. In order that these words not be grouped into a highly ungrammatical sentence, the words were read at the rate of one word per 1.5 s. Here 1.5 s exceeds the proximity constraint for the grouping process that creates the sense that neighboring words belong together. In this way, the participants experienced the list only as a temporal succession of separate events rather than as an unfolding group (sentence).

[2]Total body mass is the sum of fat free mass and fat mass. Fat mass in the human body is distinguished in terms of whether it is visceral and hidden, surrounding body organs, or subcutaneous and visible. Fat free mass is composed of bone, organs, muscle, connective tissue, etc – everything that is not fat. It provides a scientifically useful measure of body size, in so far as it is influenced by biological variables such as age and gender, but not by the sociological variables that influence fat mass such as socioeconomic status, education, and zip code. For this reason, fat free mass and not total mass is used in regression analyses of body size.