My Technique is 20% Faster: Problems with Reports of Speed Improvements in HCI

Pierre Dragicevic
INRIA, France

In human-computer interaction (HCI), speed improvements are often used as an informal measure of effect size, with statements such as *"technique A was about 20% faster than technique B"*. Such statements are however surrounded by confusion. A quick analysis of the literature reveals that half of the time, the term *"% faster"* actually refers to a percent change in task completion time, while one third of the time it correctly refers to a percent change in speed. The rest of the time, the meaning is unknown or the calculation is wrong. We explain why percent changes are inherently confusing, and propose to focus effect size discussions on *ratios*, or on measures of *percent difference* where the normalizer is the mean of the compared values. When percent changes need to be reported, it is suggested to focus the discussion on improvements in *task completion times* rather than speed.

I. INTRODUCTION

Taken broadly, an effect size is *"the amount of something that might be of interest"* [1]. What is ultimately of interest to HCI researchers is often how interaction techniques differ in their performance, e.g., their differences in average error rates or task completion times.

Many authors use percent changes in speed as an informal measure of effect size, using statements such as *"our results show that technique A was 20% faster than technique B"*. However, such statements are confusing and should preferably be avoided for two reasons: first, percent changes are counter-intuitive. Second, percent changes in *speed* are often conflated with percent changes in *time*, leading to inconsistent and ambiguous reports.

II. PERCENT CHANGES

A *percent change* (or percentage change) is the ratio between a change in a quantity and the quantity's initial value, expressed as a percentage. That is, if the value of a quantity (e.g., the price of an item) changes from v_1 to v_2, then its percent change is:

$$p = 100 \times \frac{v_2 - v_1}{v_1} \quad (1)$$

v_2 and v_1 do not necessarily have to reflect actual changes over time and can be, e.g., the prices of an item in two different stores. Either way, one computes p and if p is found to be positive, then one can say that *"v has increased by p%"* or that *"v_2 is p% more than v_1"*. If p turns out negative, then one would say that *"v has decreased by $(-p)$%"* or that *"v_2 is $(-p)$% less than v_1"*.

Compared to an absolute change $v_2 - v_1$, a percent change has the advantage of being unitless. Percent changes are widely used in disciplines such as medicine [9] and economics [6]. But one difficulty with percent changes lies in their asymmetry, leading to oddities. For example, saying that A is 25% more expensive than B is equivalent to saying that B costs 20% less than A.

The term v_1 in Equation 1 is called the *baseline* [9], or alternatively, the *base* [5] or the *reference value* [3]. Choosing the other value as the baseline, i.e, swapping v_1 and v_2, can dramatically change the value of p. Although the choice of a baseline is obvious when measurements have been taken at different points in time — as it is typically the case in medicine or economics — it is less so when measurements have been taken in no particular order — as it is typically the case in HCI.

Baselines are so confusing that they have been often used as a means of abusing and misleading people [5, pp. 109–112]. Although HCI articles always use phrasings that clearly specify which is the baseline, the legitimacy of chosen baselines remains dubious. The convention in HCI is to report improvements of "newer" over "older" techniques. But the limits of this convention become apparent when two novel designs or two old designs need to be compared. In medicine, we want to know if a patient got better after a treatment, and reporting percent changes is sensible. But in HCI, we simply want to compare techniques with each other, and the notion of baseline is most often artificial and unnecessary.

Aside from baseline issues, a drawback of percent changes is that they are not the most intuitive measure of difference one can think of. Suppose you are proposed a special fare for a hotel room and can pay \$100 instead of \$300. Will you think of it as a 67% discount? More likely, you will consider that you will pay only 1/3 of the price (a ratio), or possibly that you will save \$200 (an absolute difference). Or take any bar chart and try to estimate percent changes. This is hard because percent changes do not directly map to the way we naturally think of differences and proportions. Knowing this, it is hard to understand why they are so popular.

But why not take percent differences in speed for what they are, i.e, rough estimates of how impressive new techniques are? Cannot we adopt them as a convenient HCI standard and agree on the meaning of percentages, e.g., *"50% faster"* than the best technique would be revolutionary whereas *"5% faster"* would too incremental for publication? Unfortunately not, as we will now see.

III. CHANGES IN SPEED VS. TIME

The most serious issue with percent changes in speed is the easy confusion between speed and time. If technique A takes on average 20% less time than technique B, it does *not* follow that technique A is 20% faster than technique B. Speed is a different measure from time.

Suppose we are not comparing techniques but cars. We have different pilots drive car A and car B on a 10-km circuit (the *task*), and we measure how long they take to cover this distance (*task completion time*). We find that on average, it takes them 125 seconds with car A, but 200 seconds with car B. If we take B as the baseline, the percent change is $p = 100 \times (125 - 200)/200 = -37.5\%$, meaning A takes on average 37.5% less time than B to cover this distance (to complete this task).

But how do the cars differ in speed? The average speed of car A is the distance it covered divided by the time it took on average, or $10/125 = 0.08$ km/s (or 288 km/h), whereas the average speed of car B is $10/200 = 0.05$ km/s (or 180 km/h). The speed improvement of A over B is therefore $p = 100 \times (0.08 - 0.05)/0.05 = 60\%$. So car A takes 37.5% *less time* than car B to cover the same distance, but it is 60% *faster* than car B.

There is no reason why techniques in user studies should be treated differently: speed always refers to a quantity of something per unit of time. The purpose of a car is to cover distances, the purpose of a technique is to carry out tasks. Therefore, the speed of a technique should refer to a number of tasks (or pieces of task) carried out per unit of time. Actually, it can be verified that the choice of units does not affect p. For cars it could be kilometers per hour, meters per minutes, mph, etc. In particular, one can take the inverse of travel times and ignore distances: $p = 100 \times (1/125 - 1/200)/(1/200) = 60\%$. The general formula is:

$$p_{speed} = 100 \times \frac{1/t_2 - 1/t_1}{1/t_1} \tag{2}$$

Or after simplification:

$$p_{speed} = 100 \times \frac{t_1 - t_2}{t_2} \tag{3}$$

Equation 3 can also be used to compute the percent change in speed of interaction techniques, t_1 being the average task completion time for the baseline technique, and t_2 the average task completion time for the new technique. One could think of it as applying Equation 1 to the average number of tasks carried out per unit of time, but the actual units do not matter for computing p.

Now compare this with the original formula for computing the percent change in task completion time:

$$p_{time} = 100 \times \frac{t_2 - t_1}{t_1} \tag{4}$$

We can see that Equation 3 is the same as what we would have obtained by swapping t_1 and t_2 in Equation 4, i.e., by switching the baseline. To summarize:

- Saying that A takes $x\%$ less time than B is ***not*** the same as saying that A is $x\%$ faster than B.
- Saying that B takes $x\%$ more time than A is the same as saying that A is $x\%$ faster than B.
- Saying that A takes $y\%$ less time than B is the same as saying that B is $y\%$ slower than A.

The following article excerpt shows how the equivalence between *"% more time"* and *"% faster"* can be exploited in a discussion:

"Results show that the selected visual augmentation caused users to believe it had a duration equal to that of a progress bar 11% longer in actual time. In other words, visually augmented progress bars could be used to make processes appear 11% faster, when in reality, their duration remains unchanged." [4, p. 4]

So where is the problem with percent changes in speed? Can't we all simply follow those mathematical definitions? Again, the real problem is that these measures are counter-intuitive. The math may be correct, it remains difficult for readers to quickly grasp the meaning of statements like *"A takes 10% less time than B"* or *"B is 10% slower than A"*, and figure out whether there is a difference between them. Not only readers are confused but also a number of authors, as we will now see.

IV. LITERATURE ANALYSIS

We searched the term *"% faster"* in 10 conference proceedings: the last three proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI' 09-11), the last five proceedings of the IEEE Symposium on Information Visualization (InfoVis' 07-11), and the last two proceedings of the ACM Symposium on User Interface Software and Technology (UIST' 10-11).

The term appeared 89 times in 43 different papers. We used the reported descriptive statistics to determine what the term was referring to. Among all 89 occurrences:

- **≈50%** of the time (n=43), the term *"% faster"* referred to a **percent change in task completion time**, i.e., Equation 4 was used instead of Equation 3.
- **≈30%** of the time (n=28), the term correctly referred to a **percent change in speed**:
  - ≈20% of the time (n=19), the measures of interest were task completion times, and Equation 3 was used to obtain the corresponding percent change in speed.
  - ≈10% of the time (n=9), the measures of interest were speeds (e.g., words per minute), and the percent change was directly obtained from Equation 1.

- ≈**20%** of the time (n=18), the meaning was **undetermined**:
  - ≈15% of the time (n=15), the values from which the percent change was computed were not reported, in either numerical or graphical form.
  - ≈5% of the time (n=3), the percent changes were wrong, i.e., their numerical value did not match the result of any formula.

This quick analysis reveals the confusion that surrounds the term *"% faster"* in HCI, except in rare cases where the dependent variable is already a speed (e.g., words per minute). A reader can easily resolve ambiguities when the values t_1 and t_2 are provided in the same sentence or in a table nearby. However, percent changes in speed are often used to feed discussions that are remote from the descriptive statistics, including in introductions, conclusions and abstracts.

The values t_1 and t_2 are sometimes provided in graphical form only, e.g., bar charts. Although it was tedious to extract numerical values from these charts in the context of this survey, in practice charts are enough for the reader to get a clear idea of effect sizes and whenever they are provided, percent changes can be safely ignored.

Ambiguities surrounding percent changes are especially problematic when the values from which they are computed are not provided in any form (15% of the time). This includes expeditious statistics reports (especially in short papers), citations of results from previous work, and percent changes reflecting comparisons with previous work. This suggests that percent changes in speed are believed to be adequate for quickly summarizing results, but our analysis provides clear evidence to the contrary.

We know turn to two alternatives: ratios and symmetric measures of percent differences.

V. RATIOS

Physical measurements with a true zero point such as distances and times are sometimes called *ratio scales*. A natural way of comparing several such measurements is by thinking in terms of *ratios*. In the previous example, the time it took for car A to cover the circuit length was $125/200 = 0.625$ times the time it took for car B. So one could say that on average, it took for car A less than 2/3 of the time it took for car B. It is easy to build a mental picture of the corresponding bar chart.

Ratios are asymmetric but switching baselines only requires taking the inverse of the ratio. For example, saying that *"it took car A less than 2/3 of the time it took car B"* is the same as saying that *"it took car B more than 3/2 of the time it took car A"*. Time ratios and speed ratios are also related by an inverse relationship. For example the two previous statements are equivalent to *"the speed of car A was more than 3/2 the speed of car B"*.

One may want to multiply these ratios by 100 to turn them into percentages and say, e.g., *"the speed of car A was 160% the speed of car B"*, or *"the speed of car B was 62.5% the speed of car A"*. But these percentages run the risk of being confused with percent changes. Percent changes are easy to spot because they are followed by comparatives (e.g., *"% faster"*) or trend nouns (e.g., *"% increase"*). To minimize the risk of confusion, these should be avoided when reporting percent ratios.

Equations 1–4 can be rewritten so that only ratios appear, so it is possible to convert between ratios and percent changes without knowing v_1 and v_2 (below, replace a and b with *"the time for A/B"* or *"the speed of A/B"*):

- Saying that a is x% higher than b is the same as saying that a is $(x + 100)$% of b.
- Saying that a is x% lower than b is the same as saying that a is $(100 - x)$% of b.
- Saying that a is x% of b if $x \geq 100$ is the same as saying that a is $(x - 100)$% higher than b.
- Saying that a is x% of b if $x \leq 100$ is the same as saying that a is $(100 - x)$% lower than b.

Note that these conversions are simple mathematically but not necessarily cognitively, so focusing a discussion on ratios will be easier for readers than if all percent changes need to be mentally converted to ratios. Unfortunately, apart from special cases (e.g., when a technique can be said to be "twice as fast"), reporting ratios requires phrasings that are generally longer and maybe more awkward than reporting percent changes. It seems that simple concepts do not always map to simple language, while confusing concepts sometimes do.

VI. PERCENT DIFFERENCES

When discussing differences, one may prefer a measure of effect size that really captures a difference, i.e, that is equal to zero when the two values are equal. This can be achieved using *percent differences*. Several formulas exist for computing percent differences [6]. One of them – and perhaps the worst – is Equation 1. In fact, when v_1 and v_2 have no clear chronological order, "percent difference" is a more appropriate term for p than "percent change", but we chose to stick to the latter to avoid ambiguities, since there are many ways of computing percent differences.

In physics measurements [7], a percent difference generally refers to the following measure, which uses the mean value as the normalizer instead of the initial value:

$$p = 100 \times \frac{v_2 - v_1}{\frac{1}{2}(v_1 + v_2)} \tag{5}$$

This measure has the advantage of being symmetric. If we take car B as the baseline, then $p = 100 \times (125 - 200)/\frac{1}{2}(125 + 200) = -46.2\%$. If we take car A as the baseline instead, then $p = 46.2\%$. Switching baselines only changes the sign of p, so it is sufficient to report the direction of the effect and its size $|p|$, without having to specify a baseline. For example, one could say that *"car A was faster than car B, with a difference of* 46.2%*"*.

Now if we consider car speeds instead of travel times and keep car B as the baseline, we get a percent difference of $p = 100 \times (10/125 - 10/200)/\frac{1}{2}(10/125 + 10/200) = 46.2\%$. It can be verified that replacing times with speeds has the same effect as switching baselines: it only changes the sign of p. Therefore, when we reported a 46.2% difference between the cars, we did not need to specify whether we were talking about travel times or speeds.

Here are two imaginary examples of use in HCI:

Example 1. *"Technique A was found to be faster than technique B overall, but there was a clear interaction with target size. For large targets (W = 40 mm), the percent difference was only 2.5%, while it was as much as 50% for small targets (W = 1 mm)."*

Example 2. *"The superiority of technique A over technique B was consistent with previous studies but we found a larger difference. Table 1 summarizes results from previous studies. The largest previously reported difference was 11% (Smith et al.), while ours was as much as 33%."*

Since several formulas are possible for computing percent differences (including using the max of v_1 and v_2 as the normalizer, the min, the geometric means, etc. [6]), it is necessary to specify the formula used, for example in a footnote: *"all percent differences were computed using* $p = 100(v_2 - v_1)/\frac{1}{2}(v_1 + v_2)$ *in order to get symmetric measures of relative differences"*.

Arguably, the ambiguity surrounding the term *percent difference* is a drawback of this approach. Also, as with percent changes, it is difficult to mentally switch between percent differences and simple graphical representations such as bar charts. However, the symmetric nature of Equation 5 makes it a more elegant, more convenient and less confusing measure than percent change.

VII. CONCLUSION

Although speed improvements are a common measure of effect size in HCI, they confuse readers and authors alike. The term "% faster" most often refers to a percent change in task completion time, and sometimes it correctly refers to a percent change in speed. Related terms such as "% slower" likely suffer from the same problems.

Confusions stem from the counter-intuitive nature of percent changes, with odd behaviors when switching baselines and switching between times and speeds. This suggests that percent changes should be best avoided. They may be useful in economics or medicine where measurements are ordered, but most user studies in HCI involve comparing techniques independently from any notion of chronology, so percent changes are not needed.

There is still a need for reporting effect sizes in a compact numerical format, in order to assess how "impressive" new results are or for discussing previous work. We suggest to use *ratios* as an alternative. The concept of ratio is easy to grasp, directly maps to graphical representations, and one only needs to take the inverse of a ratio for switching baselines or for switching between times and speeds. We also propose to adopt the measure of *percent difference* used in physics, where the normalizer is the mean between the two compared values. This measure only requires changing sign for switching baselines or for switching between times and speeds.

Other measures of effect size exist, among which are *standardized effect sizes* such as Cohen's d that normalize absolute differences based on data distributions [2]. These are useful for capturing the variability across subjects or when the units of measurement are not meaningful to readers [8]. But for discussing familiar units, non-standardized measures (such as time or speed ratios and percent differences) are easier to understand [8].

When percent changes need to be reported and the dependent variables are task completion times, it is advised to either i) focus on improvements in task completion times and avoid terms such as "% faster" or "% slower", or ii) clearly specify what is meant by these terms.

[1] Cumming, G. (2011). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis.* Multivariate Applications Series, Routledge, ISBN 9781136659188.

[2] Fritz, C. O., Morris, P. E. and Richler, J. J. (2012). *Effect size estimates: Current use, calculations, and interpretation.* Journal of Experimental Psychology: General, 141(1), 2–18.

[3] Gluck, M. (2011). *Uses and Abuses of Percentages.* Lecture notes, Department of Mathematics, University of Florida. tinyurl.com/gluck2011.

[4] Harrison, C., Yeo, Z. and Hudson, S. E. (2010). *Faster Progress Bars: Manipulating Perceived Duration with Visual Augmentations.* Proceedings of ACM CHI'2010.

[5] Huff, D. (1954–1993). *How to lie with statistics* (illust. I. Geis), Norton, New York, ISBN 0-393-31072-8.

[6] Tornqvist, L., Vartia, P. and Vartia, Y. O. (1985). *How Should Relative Changes Be Measured?* The American Statistician, Vol. 39, No. 1., pp. 43-46.

[7] Wenning, C. J. (1994–2008). *Percent Difference – Percent Error.* Student Laboratory Handbook. tinyurl.com/wenning2008.

[8] Wilkinson, L. and the Task Force on Statistical Inference. (1999). *Statistical methods in psychology journals: Guidelines and explanations.* American Psychologist. Vol. 54(8): 594–604.

[9] Zhang, L. and Han, K. (2009). *How to Analyze Change from Baseline: Absolute or Percentage Change?* D-level Essay in Statistics. Dalarna University. tinyurl.com/zhang2009.