Journal of Phonetics xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Special Issue: Emerging Data Analysis in Phonetic Sciences

On visualizing phonetic data from repeated measures experiments with multiple random effects

Stephen Politzer-Ahles [a,*], Page Piccinini [b]

[a] The Hong Kong Polytechnic University, Hong Kong
[b] Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL* Research University, 75005 Paris, France

ARTICLE INFO

Article history:
Received 27 July 2017
Received in revised form 16 April 2018
Accepted 10 May 2018
Available online xxxx

Keywords:
Data visualization
Repeated measures
Random effects
Linear mixed effects models

ABSTRACT

In recent years, phonetic sciences has hosted several debates about the best way to statistically analyze data. The main discussion has been about moving away from analyses of variance (ANOVAs) to linear mixed effects models. Mixed models have the advantage both of allowing for including all data points produced by a participant (instead of computing means for each participant) and accounting for both by-participant and by-item variance. However, plotting of data has not always followed this trend. Often researchers plot participant means and standard error (as based on the number of participants), which, while potentially representative of the data used for an ANOVA, do not match the data used for a mixed effects model. The present paper discusses the shortcomings of traditional data visualization practices, solutions to these shortcomings that have been discussed in recent years, and the special challenges that come with trying to extend these solutions to phonetic data with crossed (within-participant and within-item) designs. For each of the problems discussed, we provide examples with simulated data to demonstrate how different plotting techniques can correctly, or incorrectly, represent the underlying structure of data. Ultimately we conclude that there is no single type of plot that can show everything one needs to know about this type of data, and we advocate for an approach that involves using different types of plots throughout data analysis, and making data publicly available.

1. Introduction

The last ten years have seen many advances in statistical analyses in phonetic sciences. In inferential statistics, researchers have been using more advanced models such as linear mixed effects models (LMEMs) to test experimental hypotheses. These models give researchers the flexibility to account for several kinds of variance in the data. Predictive analytics has also become increasingly popular in phonetics, with researchers using predictive models to do things such as classify or cluster sounds given a set of phonetic features. Indeed, this special issue looks at the rise of different statistical methodologies in the field, trying to better understand which can potentially be most useful to phoneticians.

While most discussion has focused on statistical analyses, another aspect of data analysis that has begun to receive more attention is data visualization. The relationship between statistics and visualization is interesting, as the choice of a statistical test has often guided the method of visualization. For example, when using analyses of variance (ANOVAs), researchers often find the mean of a given dependent variable and then plot a representation of that mean, with a standard error based on the standard deviation, the number of participants, and the t distribution. However, this method can often obscure important information about the data, information that should inform the model used for statistical analysis.

Recently there has also been increasing awareness of the importance of considering variation when making statistical conclusions, and the value of visualizing the data rather than simply relying on dichotomous judgments (i.e., “significant” or “not significant”) based on inferential statistics. Regardless of whether a pattern is significant, it is important to be aware of things like how many participants show the pattern, particularly when making conclusions about the practical or psychological importance of a finding. Accordingly, recent years have seen the publication of several papers with valuable exhortations and recommendations about how to improve our visualization practices to show how reliable (or un-reliable) results are

* Corresponding author.
E-mail address: stephen.politzerahles@polyu.edu.hk (S. Politzer-Ahles).

https://doi.org/10.1016/j.wocn.2018.05.002
0095-4470/© 2018 Published by Elsevier Ltd.

across participants (Rousselet, Foxe, & Bolam, 2016; Weissgerber, Milic, Winham, & Garovic, 2015). However, data from phonetics experiments (as well as other kinds of experimental psychology data, such as psycholinguistic data) often raise another problem, that of items or stimuli. In phonetics one usually wants to make general conclusions about some phenomenon – how something is realized in a given language, in a given speech context, or in a given population, etc. This requires not only generalizing beyond the participants who took part in the experiment, but also generalizing beyond the specific stimuli – words, sentences, contexts, etc. – that were used in the experiment. The need for statistical methods that allow for inferences beyond the items tested has been known for over 40 years (Clark, 1973), and in the last decade LMEMs have emerged as a popular and powerful technique to facilitate inferences both beyond the participants tested and beyond the items used (Baayen, Davidson, & Bates, 2008; Chang & Lane, 2016; Judd, Westfall, & Kenny, 2012). Are there also visualization techniques to facilitate such inferences? Just as Rousselet and colleagues (2016) and Weissgerber and colleagues (2015) recommend that simple *t*-tests or ANOVAs should be supplemented by visualizations showing how a pattern varies across participants, it is also important for LMEMs to be supplemented by visualizations showing how a pattern varies across both participants and items (and whatever other relevant repeated-measures factors there are). Here we take up the question of whether or not this is possible.

In this paper we summarize some arguments that have been made to advocate for better data visualization practices, and go on to illustrate why it is challenging to carry these practices out effectively when it comes to common experimental designs in phonetics and psycholinguistics. After illustrating why it is not possible to simply plot all the data, we review some potential solutions and their limitations. We end by giving some suggestions for types of visualizations that show the important aspects of the data as much as possible, discussing the importance of using different types of visualizations for different purposes at different stages of data analysis and presentation, and emphasizing the importance of data availability. Through these examples we hope to illustrate that data visualization for designs typical in phonetics – experiments with repeated measures for both participants and items – is challenging and has no one-size-fits-all solution, but requires an awareness of the advantages and disadvantages of each visualization technique for each stage of data analysis.

2. Problems in data visualization

Data visualizations serve many different purposes, such as aiding steps of data analysis (such as identification of outliers), informing statistical inferences, and communicating patterns of results to others. In this section we will focus on challenges in making statistical inferences from plots; in Section 3 we will discuss other relevant functions of data visualizations.

2.1. Capturing the data distribution

A necessary first step in any analysis is to have a sense of your data's distribution. Lacking a full understanding of the distribution of your data can have implications for both data visualization and data analysis. In phonetic sciences it is common for data visualizations to show a measure of central tendency (e.g., a location parameter such as the mean) and a measure of variance or precision (e.g., error bars representing a scale parameter like standard deviation, or representing standard error). It has long been known, however, that such plots hide potentially important information about the shape of a distribution (e.g., Anscombe, 1973).

For example, imagine you have two conditions, and both conditions have the same mean and standard deviation, suggesting that they are the same. However, on closer inspection it turns out that the two conditions have very different distributions (e.g., normal and log), and as a result cannot be considered the same. See the example in Fig. 1 showing how the same data can look different depending on how it is plotted (note that the code for this and all other plots in this article is available at https://osf.io/pm82v/). In the bar plot the two datasets look the same in terms of their means and standard deviations, but boxplots and scatter plots make it clear that they have different distributions; histograms would show this difference as well.

There are many other situations in which two datasets may differ in important ways that are not revealed in a data summary that only shows a measure of central tendency and a measure of variance or precision. For instance, distributions with very different standard errors might be this way because they have different variances, or because one has a much bigger sample size, as shown in Fig. 2. Two conditions may both have the same distribution, but a non-normal one (e.g., two datasets might both follow skewed distributions like that shown for Condition 2 in the univariate scatterplots on the right-hand side of Fig. 1), in which case the mean may not be a very accurate summary of either condition's data. Skewed distributions like these are common for types of data that have a natural lower or upper bound, such as syllable durations or reaction times, neither of which can be less than zero. Because of the abovementioned limitations of plots showing simple summary statistics, recent authors have advocated the use of visualizations which show the full distribution of data (e.g., Rousselet et al., 2016; Weissgerber et al., 2015).

These are just some of the reasons why plotting raw data, or at least diligently exploring the distribution of data, is important. Thus, recent advice such as that by Rousselet and colleagues (2016) and Weissgerber and colleagues (2015), who advise (among other things) visualizing more complete distributions of data rather than just summary statistics, is not to be taken lightly. The main argument of this paper, however, is that in some situations this advice is impossible to follow; showing all the relevant information about a dataset at once is not actually possible for many research designs common in phonetics and psycholinguistics. In addition to the challenges acknowledged by these authors (for instance, that designs with many conditions to compare are difficult to show in a single visualization), there is also a fundamental challenge. In repeated-measures experiments, the structure of a dataset is more than just the raw values; the connections between data points in different conditions is just as important. Below we will illustrate why it is not possible to show all of these connections at once in designs with crossed random effects, and offer some suggestions for strategies to show as much of the key information as possible.