Chapter 2
Counting things over time

In Chapter 1, we talked about counting things, and about how to use bar plots to represent the distribution of values of things that we can count. We also saw how useful it was to use the x-axis of a bar plot to represent the inherent order from the smallest to the largest numbers observed for a discrete numerical variable such as the count of letters in the spellings of words in a list like the *Hoosier Mental Lexicon*. The progression from earlier to later points in time is another natural way to order observations for any variable that can change over time. When we have access to a sequence of observations over time, we refer to the set of observations as a ***time series***. Since sound is something that necessarily unfolds in time, the data analysis methods that are used to make generalizations from time series data are very useful for analyzing the sounds of languages. Let's take a look at these methods, starting with the types of pictures that we can draw to represent time series data.

2.1. Time plots

One common way to represent time series data is through a time plot. These show how a variable changes over time by plotting the recorded value of an observation against the time for which it was measured. Time is almost always plotted on the horizontal axis, and the time interval between records can vary enormously from one time series to another. For example, Lonnie Thompson and other climatologists who measure variables such as the amount of dust and pollen sandwiched between annual layers of snow in ice cores drilled into glaciers and snowcaps make time plots that show data year by year, century by century, millennium by millennium, and even for longer time intervals, depending on the questions that they are asking in any given study.

For some time series, the time interval between records can even vary from one record to the next, as shown in Fig. 2.1. This time plot shows estimates of the world's human population at various time points over the last 2000 years. The first two data points are 1000 years apart, the next three are 500 years apart, and so on, with shorter and shorter intervals until the one year separating at the estimates for the world's human population in 1999 and 2000. The varying intervals between successive records reflect the amount of information that we have to make these estimates. While official population census records today often miss people in counting, we have no official population census records for any area of the world before very recently. Time plots for variables that historical linguists use in reconstructing language change over generations of speakers also often have highly irregular intervals between data points, particularly if the observations are based on analyzing whatever texts happen to be preserved from the time periods of interest.

On the other hand, many of the most common types of time series data for analyzing speech have fixed time intervals between the records, as illustrated in Fig. 2.2. The upper graph in this figure is a time plot of sound pressure level over a half second interval from the beginning of the sentence *Are you going home?* that Peter Ladefoged recorded himself saying in order to make the figure on p. 15 in his book *Vowels and consonants.* (You can get this recording from the CD at the back of this textbook or from the associated web site.) The variation in sound pressure over time is what we perceive as sound. In this case, speakers of English would perceive the vowel sound **a** that begins the word *are* in this audio wave file. The lower graph in Fig. 2.2 is a time plot of an even shorter interval in this sentence. The entire time plot in the lower graph covers just 5.6 milliseconds (a millisecond is a 1000th of a second) toward the beginning of the part of the sentence that is shown in the upper graph in the figure, specifically the interval marked off with red cursors in the upper graph. Zooming in on this very short section of the wave form lets you see the individual records as a sequence of "speckles" on the graph. In this case, there is an observation every tenth of a millisecond — i.e., every 0.0001 seconds (= 10,000th part of a second).

Since the spacing of the records in Fig. 2.2 is regular (unlike in Fig. 2.1), this means that we can specify the length of all of the time intervals between every pair of data points in the time plot by saying how many observations were made in a given unit of time. In Fig. 2.2, for example, we can say that there were 10,000 observations in 1 second. Since each observation is a "sample" of the variable being measured (here sound pressure level), this way of specifying the times for the data points in a time series is called the ***sampling frequency*** or ***sampling rate***. For recordings of speech and of sound in general, we measure sampling rate in terms of how many observations were recorded per second, and the name of this measure is ***hertz*** (abbreviated Hz).

If you look at the original recording of this sentence that you can get from the CD at the back of Peter Ladefoged's book, you'll see that there was actually a higher sampling rate than the one that we used in drawing the sound wave in Fig. 2.2. There was a record of the sound pressure level at every 0.000045 seconds, so that there 22,050 measurements per second, or a 22,050 Hz sampling rate. In recording music, an even higher sampling frequency is typically used. The standard music CD has a sampling frequency of 44,100 Hz, meaning that the sound pressure level from the microphone(s) in front of the performers was registered 44,100 times in every second of the performance.

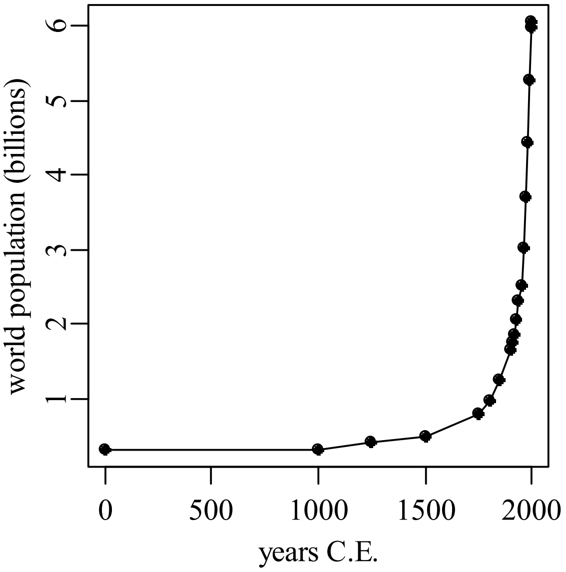

Figure 2.1. World population in billions of people, estimated for various times over the past 2000 years (see Wiesz, 2004, and the United Nations Population Division web site cited in the references).

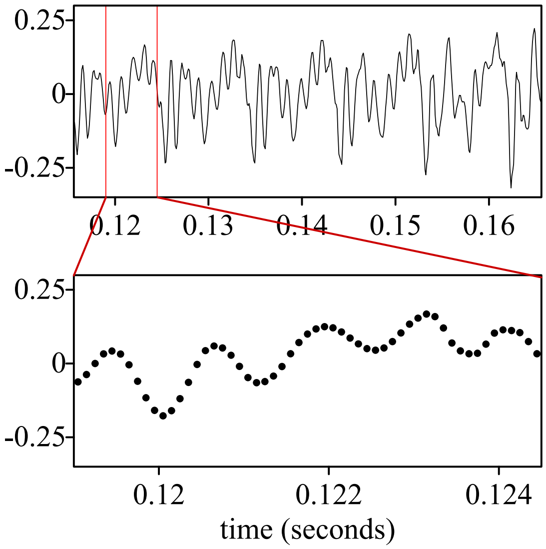

Figure 2.2. A time series plot of sound pressure over a half-second interval (top) and over a much shorter interval (bottom) taken from the beginning of the recording for Figure 2.5 in Ladefoged (2003), p. 15.

2.2. Alternative ways of counting time

Another way in which time plots can vary is that we can specify other units beside real time units along the x-axis. That is, the way that we count time on the x-axis doesn't necessarily have to be in millennia, years, days, minutes, seconds, or milliseconds. Sometimes it is more meaningful to use something else, such as the record number, as the unit of (notional) time. The time series data plotted in the two graphs in Fig. 2.3 are an example. Each of these graphs shows the average of thirteen Ohio State University students' responses in a training study where Grant McGuire tested to see how much feedback English speakers need before they can learn to reliably identify the two Arabic consonant sounds **h** and **x**. (The Arabic **h** sound is like the English **h** sound at the beginning of words such as *hair* and *holiday*. The Arabic **x** sound is like the German consonant sound **x** at the end of words such as *Dach* 'roof' — a sound which some English speakers say in the German name *Bach*.) Grant taught the labels 'h' and 'x' to each person who participated in his study, and then played many examples of the syllables **ha** and **xa** pronounced by a woman who is a native speaker of Arabic. Each time a syllable was played is a "trial" and a trial could last a shorter or longer time, because there was a pause while the person who was participating in the experiment labeled the stimulus either as an instance of 'h' or as an instance of 'x' and then got feedback about whether the label was right or wrong and what the correct label was. There were

240 such trials for each listener. Since Grant wanted to see how the amount of feedback helps listeners learn the two consonant sounds, it doesn't make sense to plot accuracy against seconds. Instead, the time unit along the x-axis is the trial number. That is, Fig 2.3 shows the responses for each trial, averaged across the thirteen people who participated as listeners in Grant's training study. If you squinch your eyes, you can (sort of) see that accuracy increases over the first 50 or 60 trials and then levels off at about 100%.

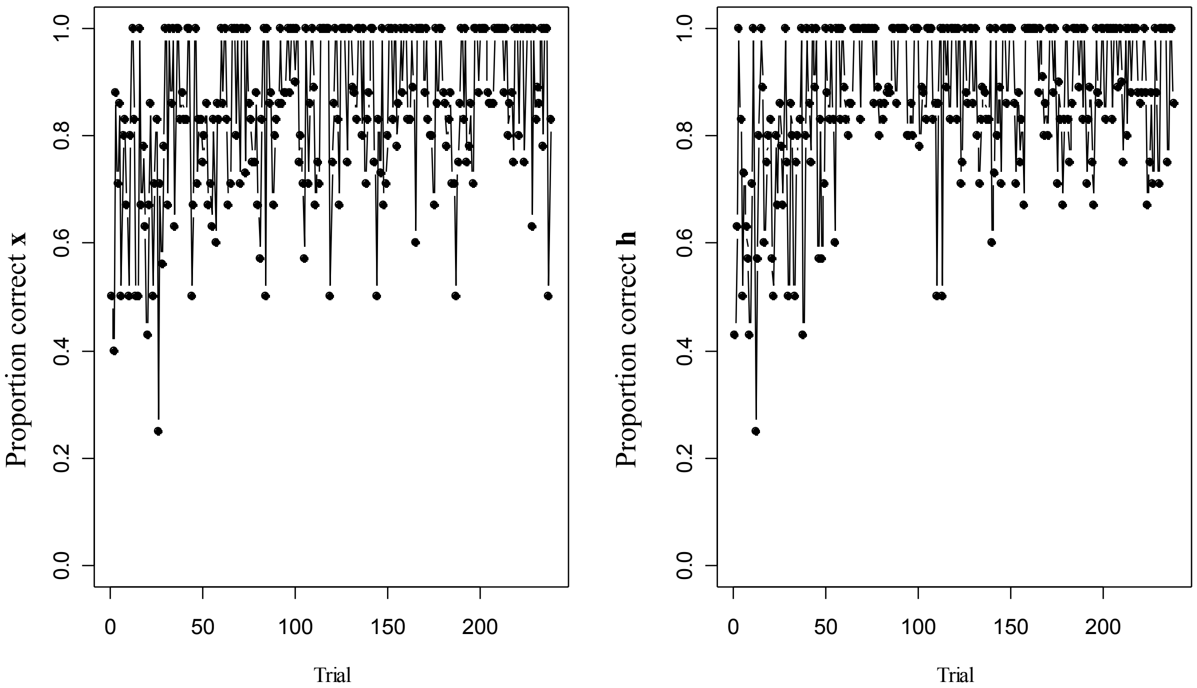

Fig. 2.3. Labeling accuracy of the Arabic consonants **h** and **x** for all training trials in McGuire (20005).

The top two graphs of Fig. 2.4 illustrate a kind of observation that you might want to plot using this kind of alternative time unit. The first of these graphs is the same time plot as in the top panel of Fig. 2.2 and the second is a time plot for another half-second interval later on in the sentence, around the vowel sound **u** in the word *you*. (The middle graph in the figure shows the time plot for the sound wave of the whole sentence, with the beginning and end of each of these two shorter time series marked off by the pairs of red or black cursors.) The five dashed vertical lines that are drawn on top of the first graph and the six dashed vertical lines that are drawn on top of the second graph demarcate the sections of sound wave that are fairly similar in shape to neighboring sections. Each of these sections that is marked off is like the "section that is repeated every one-hundredth of a second" in the figure on p. 7 in Chapter 1 of Peter Ladefoged's book. Remembering what he said there about how the "pitch of a sound depends on the rate of repetition of the sound wave," we might want to plot the repetition rate for these two sequences of five or six repeating intervals, as a first step in understanding why the pitch on the word *Are* sounds lower than the pitch on the word *you* in this sentence. The left-hand graph in Fig. 2.5 shows these two time plots. The duration of each repeating section is plotted against which section it was. This way, we can overlay the time plots for the earlier and later vowel, so that it is clear that the repeating sections are longer during the **a**. That is, you can see that the durations of the five intervals taken from the vowel **a** are all about the same length, and each of them is shorter than the analogous interval in the vowel **u**.

The right-hand graph in Fig. 2.6 then shows the corresponding frequencies, calculated by taking the inverse of the duration of each section. Remember how we calculated the 10,000 Hz sampling rate for the utterance by taking the inverse of the 0.0001 second duration of the interval between each pair of speckles in graph? It's the same relationship here between the duration of each repeating section in the time plot on the left in Fig. 2.4 and the corresponding frequency in the time plot on the right. The lower frequencies of the five repeating sections in the **a** are what make for the sensation of lower pitch there.

For the sake of comparison, we've made a time plot for the durations of such repeating sections over the whole utterance and another time plot for the corresponding sequence of frequencies. These are shown in the two graphs in Fig. 2.6. (Even though the time units here are not seconds, the time plot on the right should look familiar if you've already read Chapter 2 in Peter Ladefoged's book.)

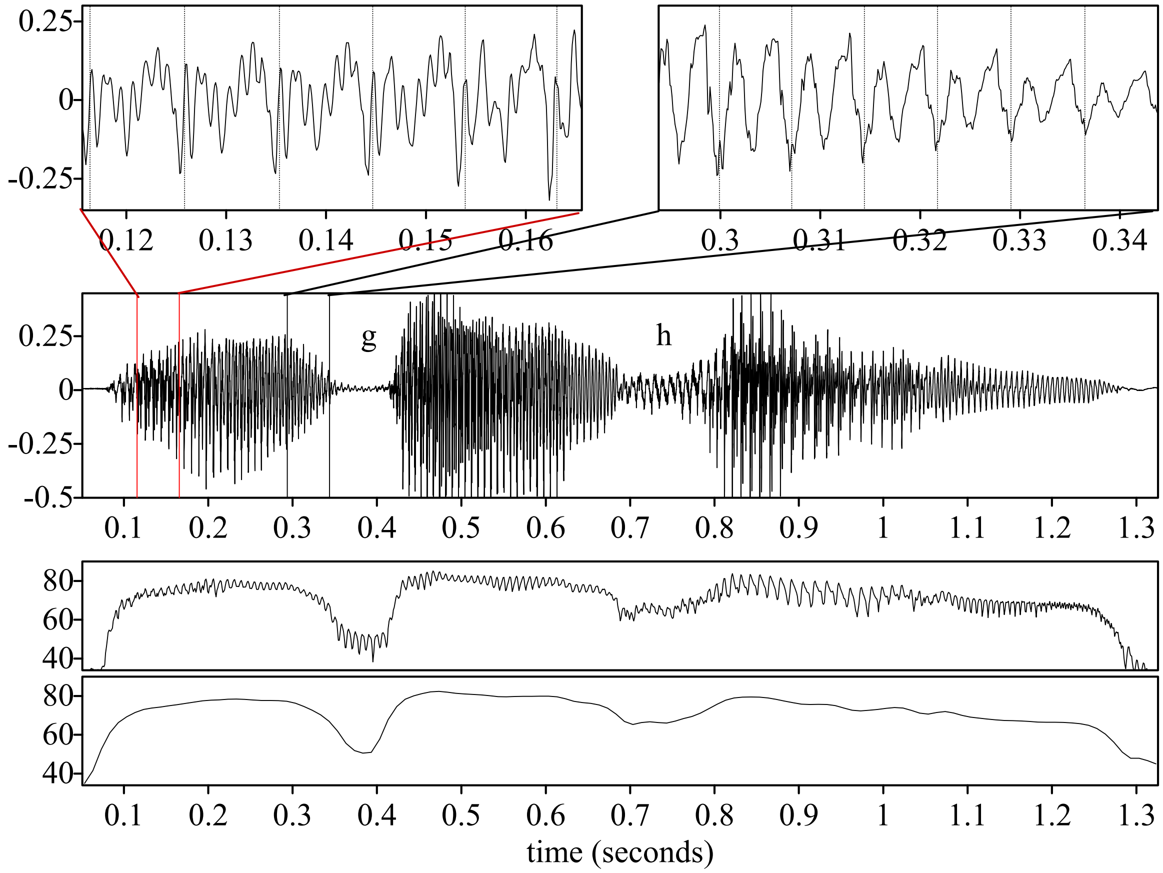

Fig. 2.4. Repeating sections marked off in the time plots for two 0.5 second intervals of the sound wave for the sentence *Are you going home?* produced by Peter Ladefoged (top two graphs) with the intervals chosen marked off in the time plot for the whole sentence (middle graph). [The first of these sections is the same as the time plot in the top graph of Fig. 2.2.] The bottom two time plots show smoothed versions of the sound wave in the middle graph, using an averaging windows of 0.001 seconds for the more jagged one and 0.01 seconds for the less jagged one.

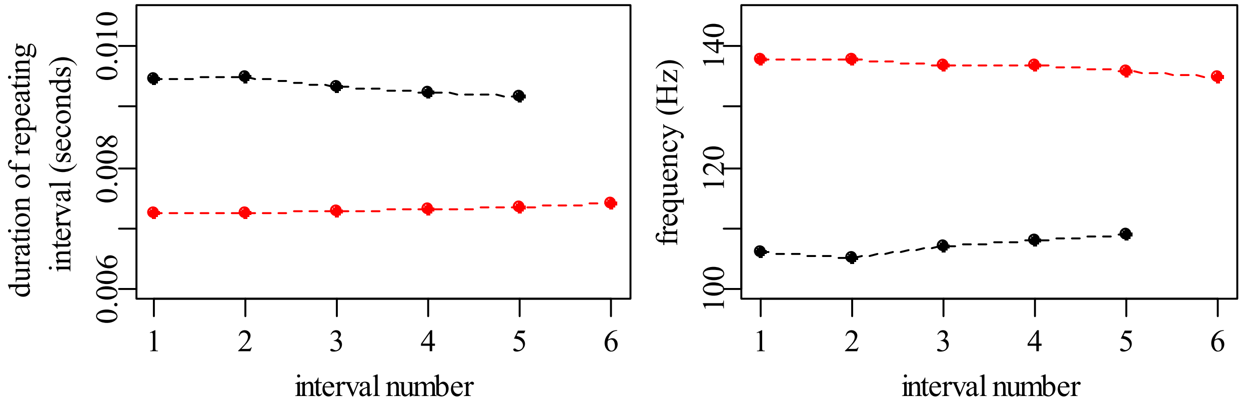

Fig. 2.5. Duration (left time plot) and frequency (right time plot) for each of the 5 or 6 repeating sections marked off by the dashed lines in the top two time plots in Fig. 2.4, with red for the **a** interval in *Are* and black for the **u** in *you*.

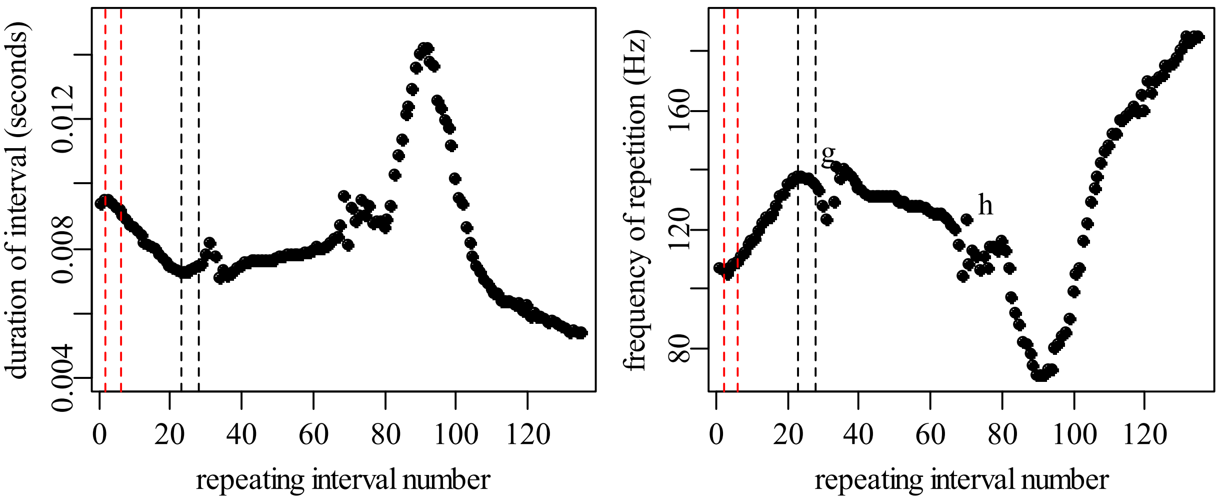

Fig. 2.6. Duration (left time plot) and frequency (right time plot) for each of the 5 or 6 repeating sections marked off by the dashed lines in the top two time plots in Fig. 2.4, with red cursors for the **a** in *Are* and black for the **u** in *you*.

2.3 Smoothing

We said above that, in their work on measures of global temperature and climate change, Lonnie Thompson and his colleagues use many different time scales, from years to millennia to even longer intervals of time. You can appreciate the reasons for this variability when you compare the time scales that you would use in predicting temperatures over the course of the day at any time of year (to decide whether to pack a sweater when going on a summer holiday trip, for example) and the time scales that a house owner might use in choosing among different varieties of peach trees to plant in the garden of the house. To get a good sense of the fluctuations in temperature from morning to evening for the city where you are going, you might consult tables of average high temperatures and low temperatures over the last few weeks. By contrast, the house owner wants to consult tables of average winter lows over the course of the last few decades, if such records are available for the area.

We can use this same idea of averaging over shorter or longer intervals of time to help interpret time plots even when the unit of time is not real time, as in Fig. 2.3. One disadvantage of plotting accuracy rates trial by trial in the way that we did there is that the pattern is very jagged. On one trial, the accuracy might be higher than on the preceding trial, but on the next, it might be lower again. Because of the large number of trials and the large variation from trial to trial, the overall trend toward better and better discrimination between the consonants is difficult to see. However, we can treat this discrete time unit "trial number" as if it were a continuous measure of time, to add up two or more trials to make longer units, as shown in Fig. 2.7. This figure shows the same data as in Fig. 2.3, but with the accuracy averaged across 10 trials. This averaging over successive trials smoothes out the jaggedness in the pattern of responses in Fig. 2.3 to better show the general learning trend. This is kind of the opposite of what we did in zooming in on the very short interval in Fig. 2.2 to be able to see the individual records over time. Here we are zooming out by averaging across multiple trials to remove the detail about trial-by-trial variation and make the general improvement over all trials clear.

The lower two time plots in Fig. 2.4 illustrate the same idea of using averages to smooth over individual data points in a time series, applied to make a rough measure of loudness changes over the sentence plotted in the middle graph of the figure. (These loudness changes are part of what makes for the sense of the rhythm of the utterance.) Recall from Peter Ladefoged's discussion of amplitude on pp. 7-8 of his book that the sensation of loudness depends on the size of the pressure variation over time. But pressure variation can be negative as well as positive, as you can see by the y-axis labels in Fig. 2.2 and the middle graph of Fig. 2.4. That is, sound is what your perceive when some event in the world (like a clap of thunder or Peter Ladefoged's vocal folds hitting each other) causes pressure to alternate between being higher than the resting air pressure and lower than the resting air pressure. Each of the repeating sections that are marked off in the upper two time plots in Fig. 2.4 is the sound of one hitting together of the vocal folds. Similarly, the large spikes that you see shooting up and down from the center point of the middle time plot is the repeating interval that results when the vocal folds smash into each other over and over again.

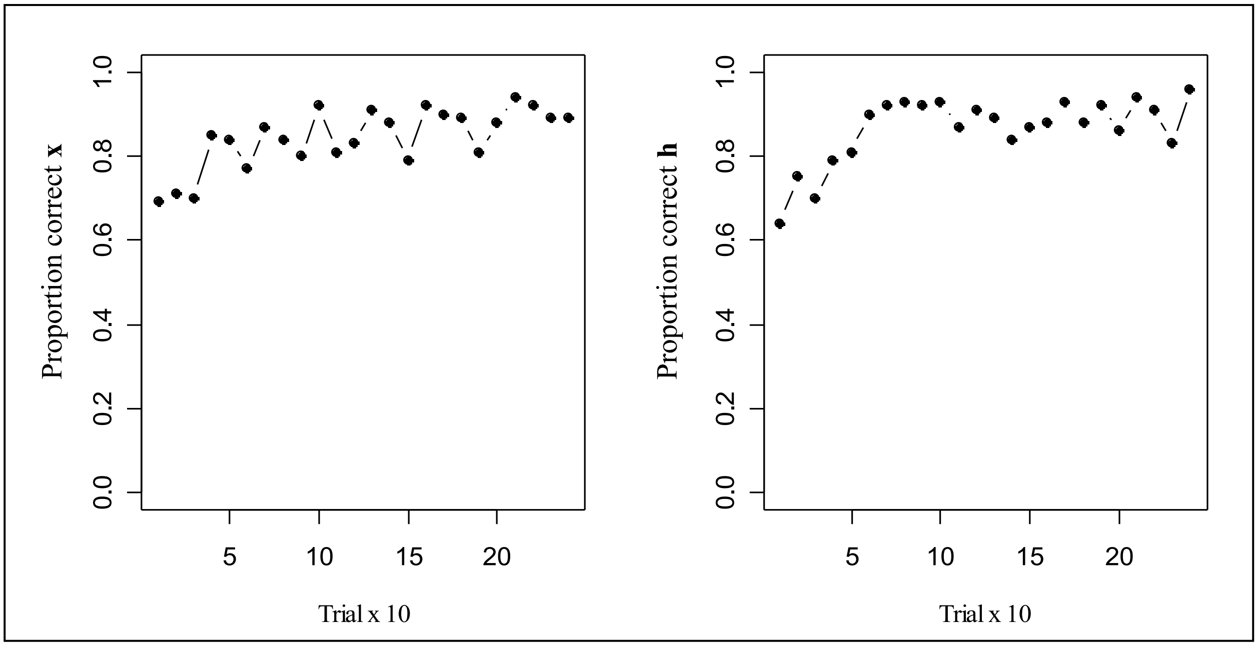

Fig. 2.7. The same data as in Fig. 2.3, but with accuracies averaged across sequential groups of 10 trials.

In order to compare the relative loudness of different sections of the sound wave, to figure out the rhythm of alternating louder and softer parts for the different syllables in the sentence, we need a way to average over the variation in pressure in both directions. One way to do that is by taking the pressure level recorded at each time point (every 10,000th of a second for this sampling rate), calculating the square of that level (to turn the negative numbers into positive numbers), and then averaging over all of these squared values in some time interval. The time plot just underneath the sound wave for the whole sentence shows the amplitude averaged in this way over a sequence of 10 samples (i.e., a smoothing window of 0.001 seconds). The time plot at the bottom of the figure shows the same thing except that here the averaging is done over 100 samples (i.e., a smoothing window of 0.01 seconds). You can see that the shorter smoothing window is more jagged, although both of these time plots are much less jagged than the raw sound wave.

The left-hand time plot in Fig. 2.8 is a similar application of averaging to smooth the frequency pattern that we showed in the right-hand graph of Fig. 2.6. Each data point in this graph is the average frequency over a group of ten intervals surrounding the repetition interval of interest. Since there are 136 repeating intervals in the sentence as a whole, there are 13 data points plotted in this smoothed frequency contour.

The right-hand time plot in Fig. 2.8 shows a different way of smoothing the frequency contour for the same sentence. Here the original frequency values are used, but only a subset of the data points are kept. We choose the data points to keep by looking at the change in frequency from one repeating interval to the next. If there is a change in frequency of at least 5 Hz, then we keep the second data point. Otherwise we delete that data point and compare the following one. This is like taking the daily low temperature and the daily high temperature and throwing out any temperatures in between. This way of smoothing is called “stylization” and it is often used in writing programs to generate the pitch pattern in computer speech synthesis.

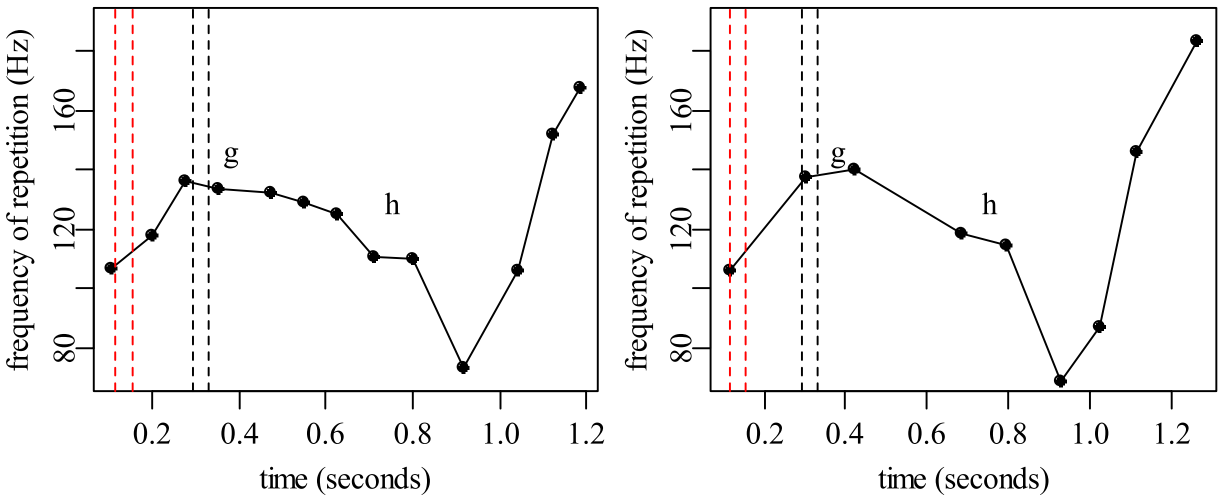

Fig. 2.8. The same data as in the frequency time plot in Fig. 2.6, but with the data smoothed by averaging across sequential groups of 10 intervals (left graph) or by throwing out data points when frequency changes slowly (right).

2.4. Summary

A time series is a variable for which the observations are ordered in time. Most ways of understanding and analyzing time series data start by making a time plot — a graph where the x-axis values are the times (or the ordinal number) of the successive observations, and the y-axis value is recorded value for the variable. Because sound is audible variation in air pressure over time, spoken words and sentences are inherently time series data, and the analysis methods used for other kinds of time series data are very useful for analyzing the sounds of languages. For example, a sound wave figure is a time plot of the varying air pressure measured with a microphone, say, and plotted over time. Another useful variable that can be plotted over time is the frequency of repetition intervals, a variable that corresponds to the sensation of pitch. (This is why Peter Ladefoged calls each time plot of frequency data in Chapter 2 of his book a "pitch track".) A time plot with jagged "noise" that obscures the trend of interest can be smoothed by averaging over successive groups of data points or by picking out data points that represent drastic changes in the variable being plotted. Both of these methods are useful in plotting aspects of the rhythms and melodies of sentences.

References

To appreciate the variable time scales used in plotting data from ice cores and the like, see:
Lonnie G. Thompson, Ellen Mosley-Thompson, M.E. Davis, P.-N. Lin, K. Henderson, and T.A. Mashiotta (2003). Tropical glacier and ice core evidence of climate change on annual to millennial time scales. *Climatic Change* 59, 137-155.

The population data in Fig. 2.1 are from:
Paul B. Weisz (2004). Basic choices and constraints on long-term energy supplies. *Physics Today*, 57 (7) [July 2004], pp. 47-52. [The table of numbers is also available from the United Nations website at: http://www.un.org/esa/population/publications/sixbillion/sixbilpart1.pdf]

The utterance depicted in Figs. 2.2, 2.4, 2.5, 2.5, and 2.8 is from the CD at the back of:
Peter Ladefoged (2003). *Vowels and consonants: An introduction to the sounds of languages*, 2nd edition. Malden, MA: Blackwell Publishing.

The training data in Figs. 2.3 and 2.7 are from:
Grant McGuire (2005). Training non-native speech categories. Unpublished manuscript, Ohio State University.