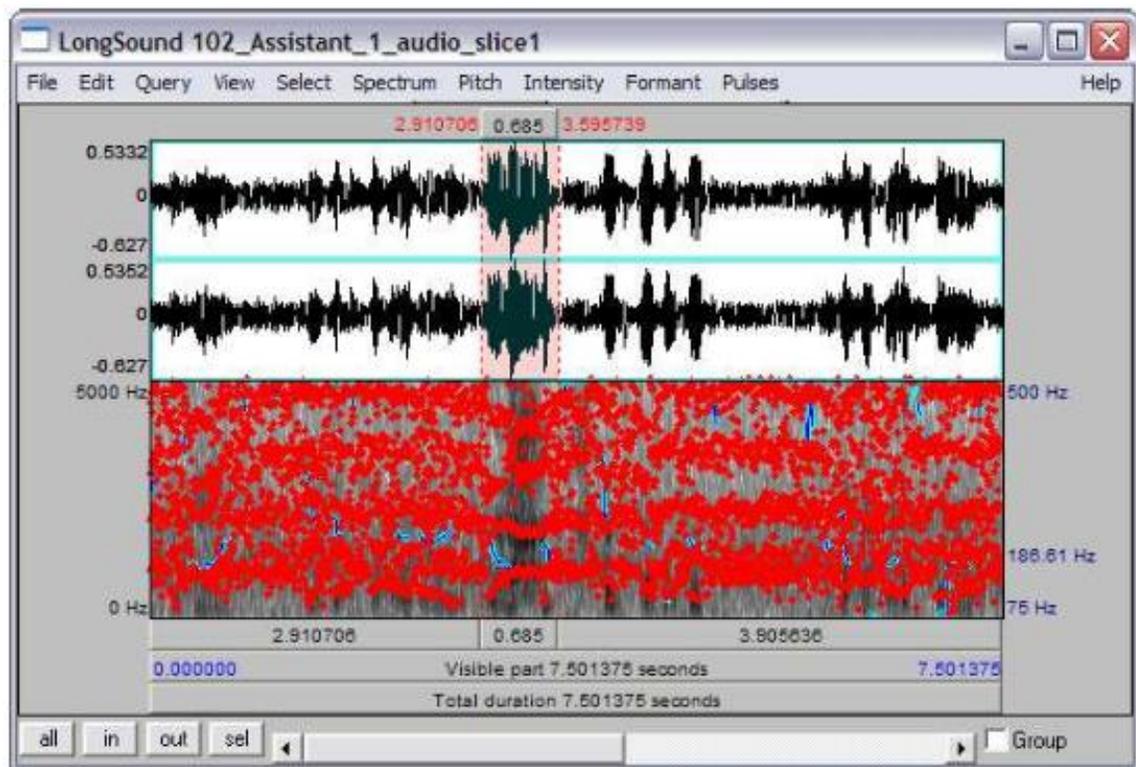


VFF Training Manual

A brief tutorial for calculating vocal fundamental frequency (VFF) using PRAAT

Christopher D. Moore
Lakeland College

Traci N. Tucker
Stanford University



© 2011 Christopher D. Moore. All rights reserved.

Table of Contents

- I. Introduction
- II. Background
- III. Example Research Design
- IV. Using PRAAT
 - Overview
 - Getting Started (Set-Up)
 - Data Analysis
- V. Recording Data
- VI. Tips
- VII. Glossary
- VIII. References

I. Introduction

This tutorial is designed for social scientists who want to learn more about calculating [vocal fundamental frequency \(VFF\)](#), the base component for vocal fundamental frequency accommodation (VFFA) (Moore 2006), using [PRAAT](#). VFFA has shown promise as a minimally-intrusive and temporally-sensitive indicator of internalized status beliefs and group structures. This manual uses an example study to illustrate how the PRAAT procedures outlined in Part III can be manually adapted to fit many experimental protocols and acoustic data.

This manual is not intended to be a general introduction to acoustics research or the PRAAT program. For information on more general uses and applications of PRAAT, please consult the beginner's manual written by Sidney Wood (<http://www.fon.hum.uva.nl/praat/>).

II. Background

Status and Vocal Fundamental Frequency

In an effort to estimate relative [social status](#), people actively search for clues designating others occupation, level of education, intelligence, apparent skill at performing specific tasks, and many other factors. These “clues” or indicators are called [status cues](#). Berger et al. (1977, 1986) determined that an individual's marshaling of status cues represents one aspect of self-presentation and serves as a basis for claiming contribution to a group's success.

Our perception of vocal pitch is related to the fundamental frequency (F_0), also called the first harmonic frequency, of the spoken sound waveform we hear. When viewed on a sound spectrograph, the human voice is revealed as a complex waveform made up of many component sound waves. Fundamental frequency (F_0) refers to the highest common factor of a periodic waveform, measured in Hz (or kHz). Put another way, the word “frequency” refers simply to how often something happens in a certain period of time; for example the average number of times a person blinks during a day. In this case, the frequency we are interested in pertains to the number of glottal vibration cycles (vocal fold

vibrations) per second. A single glottal vibration cycle consists of the opening and closing of the glottis (the opening between the vocal folds (“vocal cords”) through which air passes, located in the larynx) during speech. The sounds we hear in the human voice are the result of these vibrations. F_0 is simply the highest common frequency component shared by all of the various sinusoidal components of a complex waveform.

Research by Gregory and Gallagher (2002) locates a possible “commanding presence” vocal signature in patterns of high fundamental frequency (F_0) variability and the accommodation of others’ voices in speakers’ F_0 pattern. While the fundamental frequency between 0.0 kHz and 0.5 kHz has been identified as a critical acoustic element in communicating dominance (Gregory and Webster 1996; Gregory, Dagan, and Webster 1997), a specific mean F_0 within this range has not been identified by researchers as more central than others to the creation of a “commanding presence.”

The broader research program of Stanford Gregory and colleagues offers evidence that vocal frequency is of fundamental importance to the understanding of vocal adaptation (Gregory, et al. 2000; Gregory 1990, 1986, 1983); communication effectiveness (Gregory, Dagan, and Webster 1997; Gregory, Webster, and Huang 1993); and that interview partners adapt the frequencies of their voice to one another over the course of interaction according to their relative perceived social status (Kalkhoff and Gregory 2008; Gregory and Gallagher 2002; Gregory and Webster 1996; Gregory 1994; see also: Burgoon, et al. 1996).

III. Example Research Design

This study uses group interaction data collected from seventy-one three-person groups of undergraduate men and women who participated in a twenty-minute task-orientated team discussion (Robinson and Ten Eyck 1999). In a design adapted from Johnson (1993), each group participated in a thirty-minute task-oriented team discussion. Participants believed that they had been hired to form a temporary mock organization (an advertising team), with the goal of writing a radio commercial to recruit out-of-state students to their large Southern U.S. public university.

At the onset of the study, participants were assigned specific roles to play in the organization. In each group, two people of the same gender were each assigned to the role of "Assistant" and a single man or woman was assigned to the role of "Supervisor." To reinforce their assigned roles and the situational context, supervisors and assistants were given different badges, seated at labeled chairs in the "board room," and were asked to work from separate instruction booklets labeled either "Supervisor's Manual" or "Assistant's Manual." All experimental materials (including a sign hung over the board room door, scratch paper, and a sample radio commercial) were emblazoned with the mock company's logo. Finally, each group was given twenty minutes to create a fifteen second radio commercial. The group's discussion during this time was videotaped using three sets of microphones and 8mm video cameras (one per participant). These recordings were later converted to digital format (.avi) and archived for processing.

IV. Using PRAAT

The following description and instructions will refer to the above research design. Please note, however, that these instructions can be adapted to fit most any experimental protocol or acoustic data.

Overview

Recordings exist as separate audio files (.wav) and audio/visual ([.avi](#) or MPEG-2) files. While the analysis and coding requires only the audio files, it is advisable to reference the .avi files to verify the identity of each speaker. In order to obtain the vocal fundamental frequency (VFF) of each participant, the audio recordings are broken down into ten second increments, called slices. Within each slice are utterances from one or more participant. After identifying the source of each utterance, the VFF for each utterance may be calculated. In most instances, a single-source audio recording with an audio sample rate of at least 48.0 kHz, in 16-Bit Stereo (16.00 KB/second) is sufficient for analyzing the VFF for three individual participants seated less than seven feet from a single microphone.

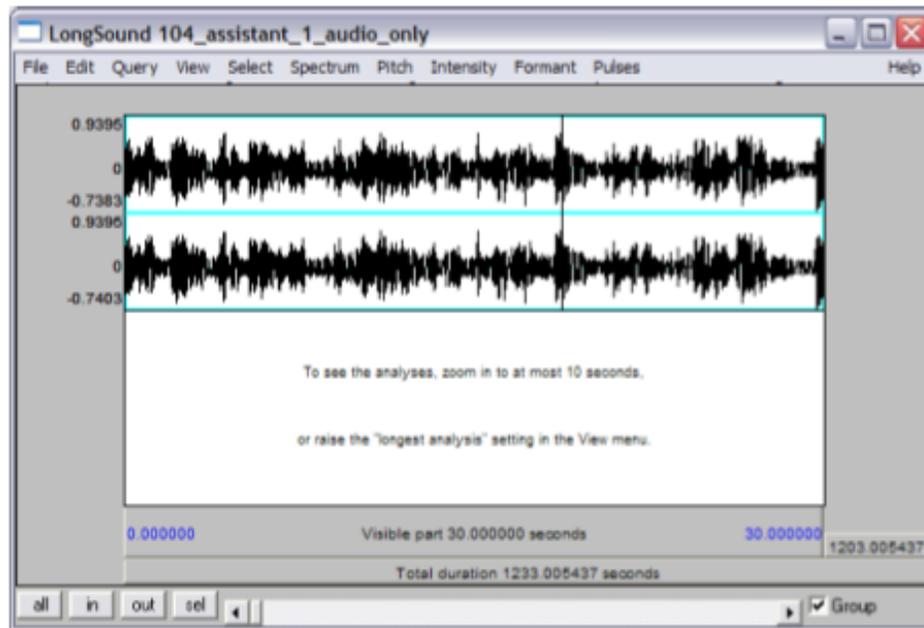
Getting Started (Set-Up)

1. Launch the PRAAT program (The PRAAT program may be downloaded for free at <http://www.fon.hum.uva.nl/praat/>)
2. PRAAT refers to the audio recordings as “[long sound files.](#)” In order to open an audio file, select “read” from the toolbar. Then choose “open long sound file” from the dropdown box.
3. A grey box will have opened on the screen. It should contain the names of all the audio files. Each name will include the group number, the subject title, and the type of data. For example, “104_assistant 1_ audio.” Select a sound file from the group on which you are currently working. The selected file will appear in the PRAAT objects box as shown below. Since you will likely be using only one sound file to analyze the VFF of all three group members, try to select the sound file that gives you the clearest recording of all three participants.



4. Click on the file name so that it is highlighted as shown in the sample image above. Then click the “view” button located on the

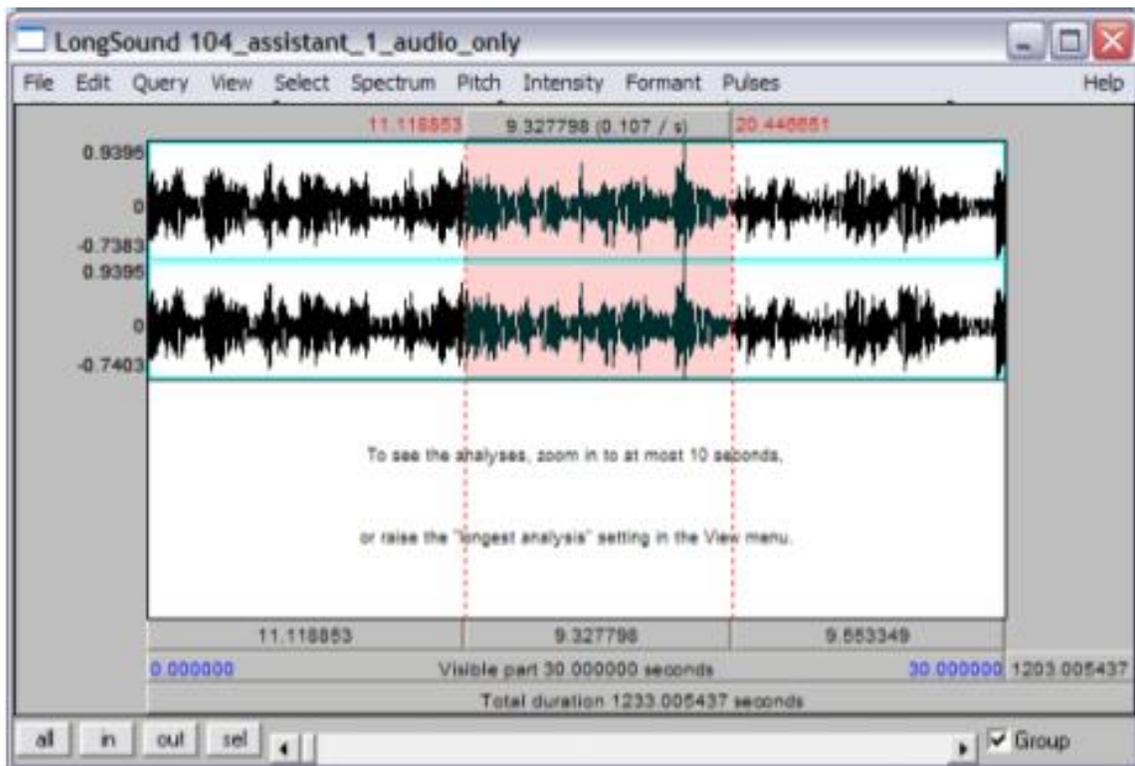
right side of the objects box. This will launch a long sound viewer as seen below. Only thirty seconds of the entire recording will be visible at any point in time.



5. Next, you will open the .avi (or MPEG-2) files that match the long sound file you have just opened in PRAAT. The video will help you to determine which subject is speaking once you begin your analysis.
6. Watch the video to determine the time at which a subject makes the first utterance. Take note of the approximate time of the utterance and pause the video.
7. In the PRAAT viewer (the box shown above), scroll through the long sound file to find the time that corresponds to the one you just noted. This is the *approximate* time of the first utterance. Because the video data and the audio data are often not perfectly synchronized, you may need to adjust your starting position in the audio file by playing it several times. To play it, click on the grey area just below the sound spectrum. To pause the sound file, press "Tab" on the keyboard. Once this starting point has been found, you are ready to begin analyzing the data.

Analyzing Data

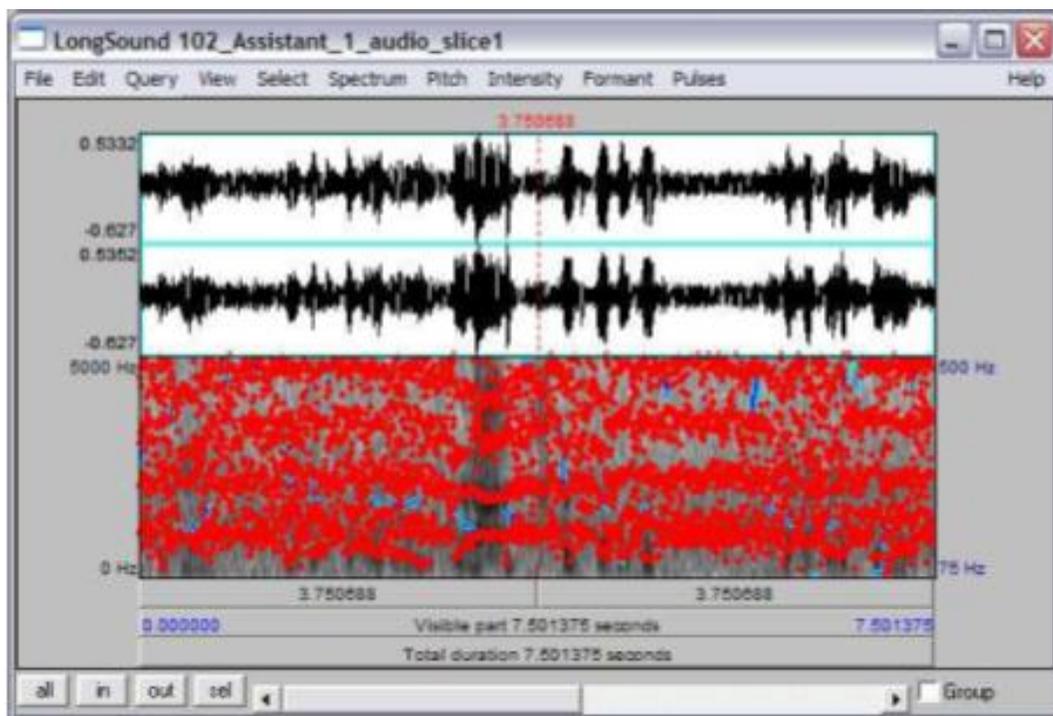
1. Note the starting time. In this example study, researchers analyzed the first three minutes of vocal data. For example, if the first utterance was made at 37 seconds, then you will stop analyzing the sound file at approximately 217 seconds (37.0 + 180.0 seconds).
2. Left click with the mouse at your determined starting point. A line will appear on your screen. Drag this line to the right until you have highlighted approximately 10 seconds of data. The highlighted area, called a [spectral slice](#), will be pink. Its length (in this case, 9.327798 seconds) will be displayed within this zone just below the sound spectrum as seen here:



3. Record in your lab book the start (e.g. 11.118853 seconds) and stop (e.g. 20.446651 seconds) times of your selection. Also record the overall time duration (e.g. 9.327798 seconds) of the slice. These numbers are found just above the sound spectrum and represent the "real time" of the entire recording. This information

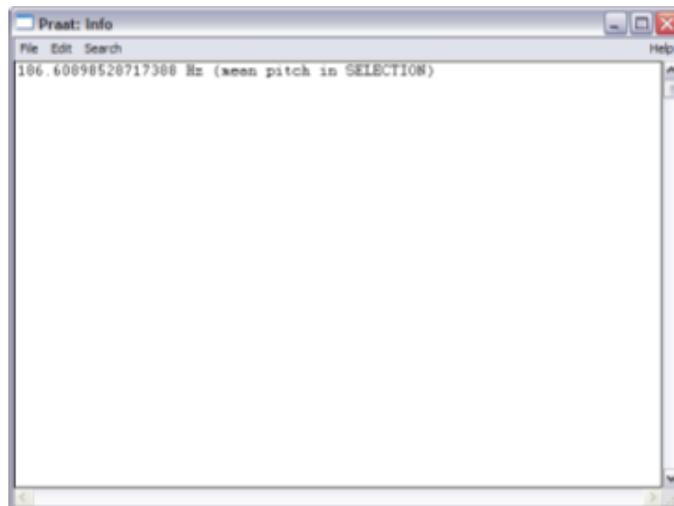
is necessary to match up the data in time-sequence later. For more information on recording data, see Part V of this manual.

4. With the slice section still highlighted, click “File” in the toolbar and select “Write sound selection to .wav file”. When saving each slice, include the group number, subject, and which slice number you are working on (e.g. 104_all_slice1, 104_all_slice2, etc.).
5. Once the slice has been saved as a new file, open it in the Pratt object viewer by again clicking “Read” on the toolbar, selecting “open long sound file”, and double clicking on the newly saved file. Once it is visible in the PRAAT objects box, highlight it and click “View”. If done properly, you should see a new box like the one shown here:



6. This box is displaying only the approximately 10 seconds of the recording that you selected in the previous procedure. Begin playing through this slice (by clicking on the grey area below the sound spectrum) to find the beginning and end of each utterance.

7. For each section where only one subject is speaking, highlight this area using the same technique that you used to highlight the original slice. This area should include the voice of only *one* participant. For example, your starting point for one such area might be when Assistant 1 begins to speak, and your stopping point might be when the Supervisor interrupts Assistant 1. Additionally, these selected areas should not be shorter than .500 seconds. Lastly, in essence, the blue lines (in the lower field) represent areas of sufficient data, or measurable F_0 . Highlighted areas should not include a gap of greater than .500 seconds between these blue lines.
8. Record the values of your highlighted section (see Part V of the manual), and determine the mean frequency of the section by clicking "Pitch" on the toolbar. Select "Get Pitch" from the dropdown menu. A new box like the one below will appear displaying the frequency. Record this value before closing the frequency box.



9. Repeat steps seven and eight for each utterance until you have analyzed the entire slice. Then you may close the slice and return to the long sound file to create your second slice. Continue the entire process until you have analyzed three minutes of data.

V. Recording Data

Since you will be recording data simultaneously for three people at a time, it is important to organize documentation not only by the group number but also by participant/assigned role. Additionally, you must make note of which one of the source (8mm tape) recordings you are using for your long wave sound file (Assistant 1, Assistant 2, or the Supervisor). In the sample data below, the recording from Assistant 1 has been used.

Entries are dated and will include the following information:

- Slice number (number used when saving slice as .wav file)
- Start time and stop time of each slice (the numbers show in red at the beginning and end of each highlighted area)
- Duration of each slice (the numbers shown in black. located just above the highlighted area)
- Start time and stop time of each highlighted area within the slice (the numbers shown in red at the beginning and end of each highlighted area)
- Duration of each highlighted area within the slice (the numbers shown in black located just above the highlighted area)
- Mean vocal frequency of each highlighted area within the slice (shown in the frequency box)

Your entries should resemble the sample data shown below:

03/24/06			
104 A1	<u>Slice 1</u>	11.118-20.446 (9.327)	
	1	2.910-3.595 (.685)	186.6090 Hz
	2	4.021-5.024 (1.003)	185.3789 Hz
	3	8.092-8.797 (.705)	199.2987 Hz
	<u>Slice 2</u>	20.603-30.600 (9.997)	
	1	5.602-6.834 (1.232)	181.8736 Hz
104 S	<u>Slice 2</u>	20.603-30.600 (9.997)	
Using A1	1	0.501-2.333 (1.832)	231.3277 Hz
Tape	2	8.345-9.213 (.868)	229.9383 Hz
	<u>Slice 3</u>	30.609-40.600 (9.991)	
	1	1.213-3.436 (2.223)	230.3874 Hz
	2	5.384-6.983 (1.599)	228.8273 Hz

Note that the durations of both the slices and the highlighted areas are always surrounded by parentheses. Highlighted areas within a slice are numbered 1, 2, 3, etc. Start and stop times are always formatted as *start time - stop time*. Vocal frequencies are designated by “Hz”, and one line is skipped between the data for each new slice. If a speaker does not speak at all during a slice, then there is no need to record any information about that slice for that specific speaker. For example, in the sample data the Supervisor does not speak during slice one.

VI. Tips

- ⇒ Because participants are seated around a circular table, when watching video from the .avi files to determine who is speaking, you will generally be able to see only two faces at a time. When the Supervisor is a different gender than the Assistants, it is helpful to use the shot that displays the faces of the two Assistants (since it is likely that it will be most difficult to differentiate between their voices).
- ⇒ The mean frequency may also be obtained by pressing F5 on the keyboard rather than selecting “Get Pitch” from the “Pitch” dropdown menu.
- ⇒ When recording data, it is helpful to have a separate piece of paper for each subject in front of you. This way, you don’t have to flip back and forth in your logbook each time a different group member begins to speak.
- ⇒ To remove items from the PRAAT objects box, highlight the items to be removed (more than one item can be selected by right clicking and dragging your mouse over the items). Then, select the “Remove” button at the bottom of the objects box.
- ⇒ If you will be stopping data analysis for the day but are in the middle of a group, it is helpful to make note of the stopping time of both the audio file and the .avi file. This way you will know where to begin when you resume data analysis. For example: .avi A1 (2:21)/ audio S (152.372).

⇒ If you are unsure about the quality of a highlighted area, place an asterik (*) beside where you recorded the information in your logbook, and make note of your reason. For example, “*background noise.”

VII. Glossary

.avi file: AVI = Audio Video Interleave, a generic Microsoft Windows-based audio/video file format

PRAAT: a computer program for speech analysis and synthesis, written by Paul Boersma and David Weknink at the Department of Phonetics of the University of Amsterdam URL: <http://www.fon.hum.uva.nl/praat/>

spectral slice: an approximately ten second selection of a voice recording

social status: position relative to that of others; social standing as related to one’s capacity to exert influence over others

status cues: indicators of relative social status characteristics such as facial features, physical proximity, vocal tone, eye gaze, etc.

vocal fundamental frequency (VFF): mean vocal pitch; first harmonic frequency, refers to the highest common factor of a periodic waveform, measured in Hz (or kHz)

.wav file: WAV = Waveform Audio File Format, a Microsoft audio file format in which spectral slices are typically saved

VIII. References

Berger, Joseph, M. Hamit Fişek, Robert Z. Norman and Morris Zelditch Jr. 1977. *Status Characteristics and Social Interaction*. Elsevier, New York.

Berger, Joseph, Murray Webster, Jr., Cecilia L. Ridgeway and Susan J. Rosenholtz. 1986. "Status Cues, Expectations, and Behaviors." *Advances in Group Processes* 3:1-21.

Berger, Joseph, Cecilia L. Ridgeway, M. Hamit Fişek, and Robert Z. Norman. 1998. "The Legitimation and Delegation of Power and Prestige Orders." *American Sociological Review* 63(3): 379-405.

Paul Boersma and [David Weenink](http://www.fon.hum.uva.nl/praat/). 2005. *Praat: doing phonetics by computer, Version 4.3.37 [Computer program]*. Retrieved December 15, 2005, from <http://www.fon.hum.uva.nl/praat/>

Burgoon, Judee K., David B. Buller and W. Gill Woodall. 1996. *Nonverbal Communication: The Unspoken Dialogue*. New York, NY: McGraw-Hill.

Ford, Rebecca and Cathryn Johnson. 1998. "The Perception of Power: Dependence and Legitimacy in Conflict." *Social Psychology Quarterly* 61:16-32.

Gregory, Stanford W., Jr. 1983. "A Quantitative Analysis of Temporal Symmetry in Microsocial Relations." *American Sociological Review* 48:129-135.

_____. 1986. "Social Psychological Implications of Voice Frequency Correlations: Analyzing Conversation Partner Adaptation by Computer." *Social Psychology Quarterly* 49:237-246.

_____. 1990. "Analysis of Fundamental Frequency Reveals Covariation in Interview Partners' Speech." *Journal of Nonverbal Behavior* 14:237-251.

_____. 1994. "Sounds of Power and Deference: Acoustic Analysis of Macro Social Constraints on Micro Interaction." *Sociological Perspectives* 37:497-526.

Gregory, Stanford W., Jr., Kelly Dagan and Stephen Webster. 1997. "Evaluating the Relation of Vocal Accommodation in Conversation Partners' Fundamental Frequencies to Perception of Communication Quality." *Journal of Nonverbal Behavior* 21:23-43.

- Gregory, Stanford W., Jr. and Timothy J. Gallagher. 2002. "Spectral Analysis of Candidates' Nonverbal Vocal Communication: Predicting U.S. Presidential Election Outcomes." *Social Psychology Quarterly* 65:298-308.
- Gregory, Stanford W., Jr., Brian E. Green, Robert M. Carrothers, Kelly Dagan and Stephen Webster. 2000. "Verifying the Primacy of Voice Fundamental Frequency in Social Accommodation." *Language and Communication* 21:37-60.
- Gregory, Stanford W. and Stephen Webster. 1996. "A Low Frequency Acoustic Signal in Voices of Interview Partners Effectively Predicts Communication Accommodation and Social Status Perceptions." *Journal of Personality and Social Psychology* 70:1231-1240.
- Gregory, Stanford W., Jr., Stephen Webster and Gary Huang. 1993. "Voice Pitch and Amplitude Convergence as a Metric of Quality in Dyadic Interviews." *Language and Communication* 13:195-217.
- Johnson, Cathryn. 1993. "Gender and Formal Authority." *Social Psychology Quarterly* 56:193-210.
- Moore, Christopher D. 2006. "Legitimacy, Status, and the Acoustic Signature of Deferential Speech." Ph.D. dissertation, Dept. of Sociology, University of Georgia.
- Kalkhoff, Will and Stanford W. Gregory, Jr. 2008. "Beyond the Issues: Nonverbal Vocal Communication, Power Rituals, and "Rope-a-Dopes" in the 2008 Presidential Debates." *Current Research in Social Psychology* 14:39-51.
- Robinson, Dawn T. and Toby Ten Eyck. 1999. "Status Ambiguity and Language-Based Negotiation of Hierarchy." Paper presented at the *Annual Meetings of the American Sociological Association* in Chicago, IL.