

# **ACOUSTIC ANALYSIS PROTOCOL**

Phonology Project Technical Report No. 4

Peter Flipsen Jr

Kristin Tjaden

Gary Weismer

Heather Karlsson

**Released: December, 1996**

*(Updated: December, 1997)*

Phonology Project, Waisman Center on Mental Retardation and Human Development,  
University of Wisconsin-Madison

Preparation of this report was supported by a research grant DC00496 from the National  
Institute on Deafness and Other Communication Disorders, National Institutes of Health,  
(Lawrence D. Shriberg, P.I.)

## TABLE OF CONTENTS

OVERVIEW .....	2
The Residual Errors Project .....	2
The Manual .....	2
The Stimuli .....	2
Analysis Overview .....	2
Figure 1. Acoustic Analysis Flowchart .....	3
GENERAL PROCEDURES .....	4
Assessing Productions .....	4
Data Storage .....	5
Basic Filename Structure .....	5
Directory Structure .....	6
Creating Directories .....	6
CSPEECH .....	7
Starting Up CSpeech .....	7
Turning CSpeech Off .....	7
Standard Analysis Conditions .....	8
Deleting/Renaming Files .....	8
ANALYSIS .....	9
Digitizing the Words .....	9
Measuring f <sub>0</sub> and Word Durations .....	12
Segmenting /s/ .....	14
Creating Moments Files .....	15
Analyzing Constriction Intervals .....	16
Creating Formant Track Files .....	18
Correcting Formant Tracks (and measuring VOTs and burst frequencies) . . . .	19
Analyzing Formant Transitions .....	22
Entering Data into Database .....	25
Reliability Measures .....	25
REFERENCES .....	27
APPENDIX A (Word Duration Criteria) .....	28
APPENDIX B (/s/ Segmentation Criteria) .....	30
APPENDIX C (Measurement Precision) .....	31

## OVERVIEW

### ***The Residual Errors Project***

This study is an attempt to find the acoustic signature for certain types of speech errors. In order to do this, both normal and disordered speakers were recruited through the local school district. A comprehensive assessment was carried out that included taping the speakers during a speaking task specifically designed for this analysis.

### ***The Manual***

This manual is intended to provide detailed descriptions of the procedures used by Phonology Project staff to extract acoustic information on speech samples produced for the Residual Errors Project. It is not intended as a manual for the use of the CSpeech software. Please consult the CSpeech program manual if you have questions about procedures that are not contained herein. It is also not intended to serve as a primer in speech science. There are a number of excellent texts in that area that you might wish to consult to learn more about what it is that you are analyzing. ***This is a how-to manual.*** Anyone with an undergraduate course in basic speech science should be able to follow the instructions given in a step-by-step manner. The task should not be overly complicated. If all else fails, consult any of the authors.

### ***The Stimuli***

The acoustics task consisted of study speakers saying words containing our target phonemes in the phrase:

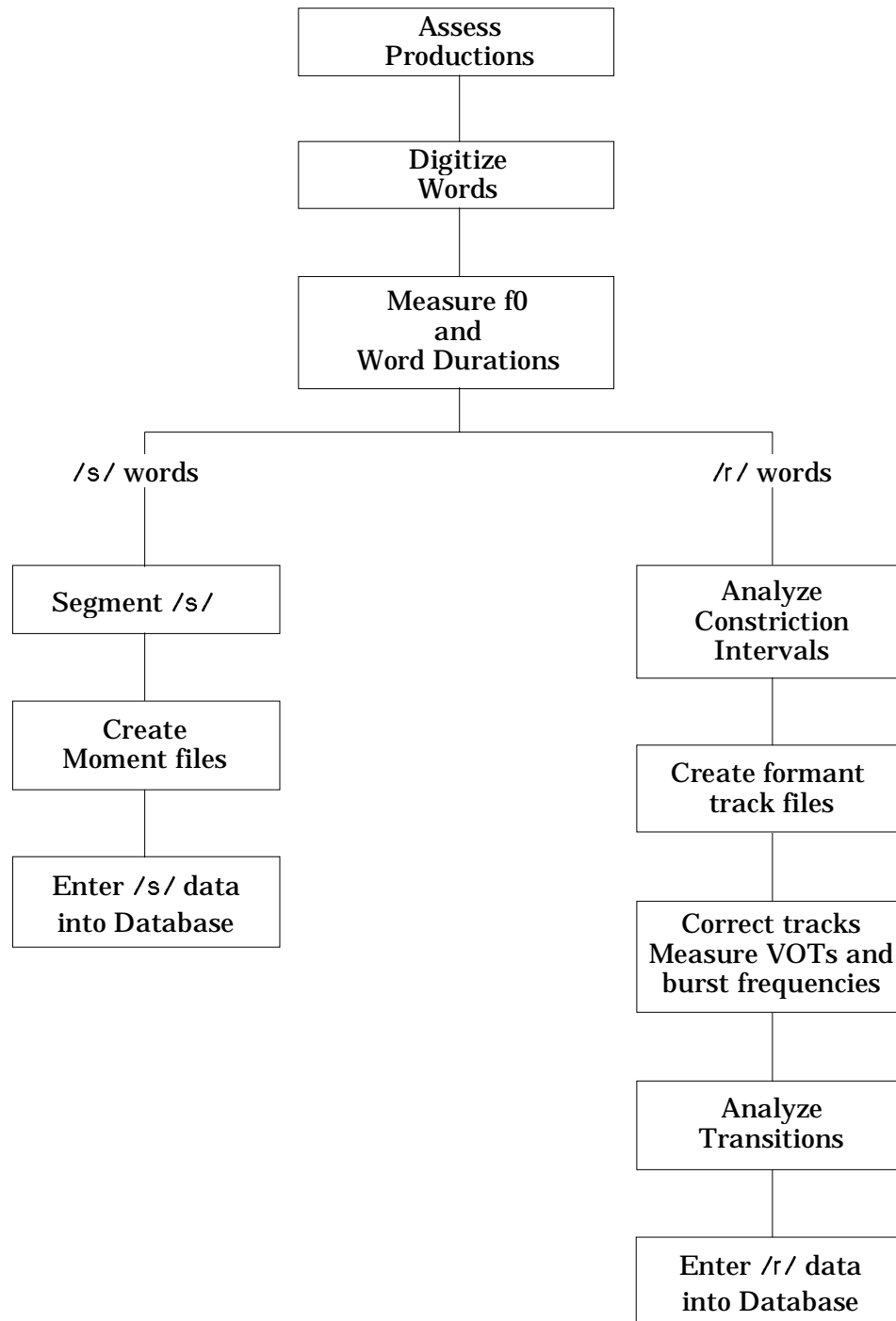
"Say \_\_\_\_\_ again"

There were 24 words and each was recorded in this phrase context 5 times yielding a total of 120 tokens.

Please note that at the beginning of each tape, the speakers were asked to prolong a series of vowels. Those vowel productions was intended to be part of a different study and should be ignored for the purposes of the current analyses.

### ***Analysis Overview***

As noted above, we are attempting to characterize the acoustic signature for certain kinds of speech errors relating to the phonemes /s/ and /r/. A series of measurements will be made on the productions of each of these phonemes. The overall procedure is outlined in *Figure 1. Acoustic Analysis Flowchart*. Each box in the figure represents a step in the analysis procedure to be carried out. The pages that follow the chart include a detailed description for each of the steps shown.

**Figure 1. Acoustic Analysis Flowchart**

## GENERAL PROCEDURES

### *Assessing Productions*

Having 120 perfectly recorded production for each speaker will be rare. All tokens must be evaluated. Each tape should be reviewed and notes made about the productions. Notes should include comments about dysfluencies or incorrect word use. Note any excessive pausing. Also note if the examiner had the speaker repeat a token and how many times this was done per token. Note whether the repetition was acceptable. We are looking for fluent productions of the targets in the specific target context. Any deviation is worth noting.

The following rules are intended to assist in determining whether a token is acceptable for analysis:

1. If speaker clearly says the wrong word (i.e. substitutes another), that token is to be **REJECTED**.
2. Productions with intrusive segments will be **REJECTED**. Examples include:

try → tried  
cry → cried                      changes syllable (open → closed)

cosine → costine  
soon → ksoon                    changes context (singl. → cluster)

cosine → cosined              spin → spinned  
assign → assigned            spoon → spooned  
pin → pinned                    skin → skinned  
rebel → rebelled              sin → sinned  
assign → gassign

kin → akin                      bird → abird (/ə/ insertion)

3. **Dysfluencies**. Any significant hesitation (pause of 250 msec or more), prolongation, or repetition of a sound, syllable or word should be **REJECTED**. The exception will be a repetition of /s/ in "say" since the result will still be a fully fluent phrase.
4. Deletions of sounds or syllables (e.g., sign/assign, gain/again) are to be **REJECTED**.

ONCE STATUS OF THE WORDS HAS BEEN DECIDED, RECORD THIS INFO ON THE FORM ENTITLED *Acoustics Task. Word Order as Presented...*

## ***Data Storage***

Data files are to be stored on VMS (not on the hard drive of the PC) because VMS is backed up every night. When you turn the PC on and login, you will be connected to VMS. Typically, project PTLAB files will be stored on drive L:\.

## ***Basic Filename Structure***

To maintain consistency, all filenames for the data files on this project should have the following structure:

*at####\_#.ext*

- all files begin with "at" (stands for acoustics task).
- this is followed by the speaker number (e.g., RE21, SD7) which will be either 3 or 4 characters long.
- then there is the letter that represents the word that was spoken (see list below).
- this is followed by the token number (ranging from 1–5).
- after the . comes the file extension (3 types possible).

*.wav*            MUST BE USED FOR ALL SOUND FILES (words & segments).

*.mmt*            is a "moments" file (created by the software).

*.fbw*            is a "formant tracks" file (created by the software).

Filename examples:

*atre21h3.wav*

Speaker RE21, 3rd token of "kicks," sound file.

*atsd7q5.fbw*

Speaker SD7, 5th token of "ride," formant track file.

—letter designations for words—

a. assign	g. crude	m. pried	u. sin
b. bird	h. kicks	n. prude	v. skin
c. burg	i. kin	o. rebel (noun)	w. soon
d. burr	j. kiss	p. rebel (verb)	x. spin
e. cosine	k. kits	q. ride	y. spoon
f. cried	l. pin	t. rude	z. tried

### ***Directory Structure***

The main directory for this project is called **ACOUSTIC**. Within that directory, there will be a subdirectory for each speaker, labeled with the speaker number (e.g., RE21, SD7).

The speaker directory will then be subdivided again into 2 subdirectories. The digitized words will be stored in a subdirectory called **WORD** and /s/ segments will be stored in a subdirectory called **S**. Thus the path for a particular word file might look like:

*L:\acoustic\re21\word\atre21h3.wav*

And the path for a particular /s/ segment file would be:

*L:\acoustic\sd7\s\sd7w3.wav*

### ***Creating Directories***

Each time we begin working with a sample from a new speaker, you will have to create a new directory and a set of subdirectories to store the files in. From the c:\:

1. type **L:\** (or the drive PTLAB is listed on) (press *Enter*).
2. type **cd\acoustic** (press *Enter*).
3. type **md (speaker #)** e.g., md RE16 (press *Enter*).
4. type **cd\(speaker #)** e.g., cd\RE16 (press *Enter*).
5. type **md word** (press *Enter*).
6. type **md s** (press *Enter*).
7. type **cd\** (press *Enter*).

## CSPEECH

### ***Starting up CSpeech***

The software program being used for this study is known as CSpeech. It was designed here at UW by Dr. Paul Milenkovic of the Dept. of Electrical Engineering. For detailed info about the program, contact him (Room 525, Waisman Center; 263-5726).

To start the program, turn on the power bar to the right of the computer. Once on, the PC will ask you to supply your username and password. Use your VMS procedure. This will take you to the DOS prompt c:\.

1. type **cd\csp** (press *Enter*).
2. type **cs** (press *Enter*).

You should now be at the CSpeech main menu. Each time you get to this point, you will be asked for some basic settings. Enter the following values:

Sampling Rate = 22  
Channels = 1                      Press *Enter* to confirm settings

### ***Turning CSpeech Off***

Whenever you wish to stop analysis and turn CSpeech off, use the following steps:

1. Save whatever you have been working on as per instructions for that particular step.
2. Use the *Esc* key as many times as necessary to return to the CSpeech Main Menu.
3. type **Q** [for Quit].
4. type **Y** [for Yes].

You were asked if you wanted to quit without saving. You should be saving as you go along so it is not necessary to save anything else at this point.

5. At the C:\csp> prompt, type **cd..** (press *Enter*).
6. type **logoff** (press *Enter*).
7. Turn off the power bar (and the amplifier if it was used).



### ***Standard Analysis Conditions***

CSpeech allows you to analyze your data in a variety of ways using a variety of filters and settings. To maintain reliability of the data, the following should be use as the mode of analysis.

1. Spectrogram BW = 500 Hz.
2. Spectrogram Range = 5500 Hz.
3. Spectrogram Floor = -72 dB.
4. Analysis Window = entire digitized form.
5. Screen in color.

Any of these can be modified to obtain additional insight into what is happening but all measurements and analysis should be conducted at the above settings. For example, on occasion the Black and White screen may allow you to see the formants a little clearer. It's okay to examine them in Black and White but track the formants using the color display. Likewise, weak formants may be made clearer by dropping the floor from -72 to -80 dB. Again, analysis should be done at -72.

**USE AS MUCH INFORMATION AS IS AVAILABLE TO MAKE THE DECISIONS.**

### ***Deleting/Renaming Files***

At several points during the analysis, it may be necessary to reject particular tokens. If this is done after the token has been digitized, note the reason for the rejection. Then all the files that go with that token must be deleted. This means any *.wav* and *.fbw* files from the **WORD** directory and any *.wav* and *.mmt* files from the **S** directory.

Once deleted any other files will have to be renamed so that there are no gaps in the number sequence. For example, if the third token is rejected, all of its files should be deleted and the 4th and 5th tokens should be renamed to be 3 and 4 respectively.

In addition, any data sheets should be modified so that the corresponding data is deleted for the deleted tokens and the values for the renamed tokens are moved to the appropriate columns.

From the CSpeech main menu:

1. type **F** (for Files).
2. type **D** (for DOS).
3. type **L:** (+ *Enter*).

- check to see that you are in the appropriate directory; if not, type **cd\** (+ *Enter*) and enter the directory path (+ *Enter*).

4. To delete a file, type **del** (+ filename + *Enter*).
5. To rename a file, type **rename** (+ filename + *Enter*).
6. Once completed, type **exit** to return to CSpeech.

NB – this can also be done from outside of CSpeech from the DOS prompt (C:\).

## ANALYSIS

Prior to conducting any analysis, please see APPENDIX C for a brief discussion on measurement precision.

### *Digitizing the Words*

Prior to digitizing a new set of words for a subject, be sure that a new set of directories has been created (see above).

In order to analyze the words using CSpeech, they must be stored in the computer in digital form. This is known as digitizing because it is a conversion of the analog signal from the audiotape into digital form in the computer's memory.

The SONY 5000EV tape recorder to the right of the PC is connected to a "sound card" which, along with CSpeech, accomplishes the analog to digital conversion.

A note about volume levels. The volume setting on the tape recorder should be optimized FOR EACH SPEAKER. Play the tape without digitizing for a minute or so and adjust the volume level so that it is at a comfortable listening level. Then digitize a few productions (as outlined below) and look at the resulting waveforms. The signal should be loud enough to create a clear distinction between speech and silence. It should not be so loud that "peak clipping" occurs (where the amplitude of the peaks is so large that the peaks would have square tops). If necessary, adjust the volume level and redigitize the tokens.

1. Startup CSpeech (see above). Turn on the amplifier (or speakers) and use headphones to monitor what you are recording.
2. type **R** (for Record).
3. Set duration = 20 sec. (press *Enter*, twice).
  - this will allow you to record 3–6 tokens at a time.
4. Advance tape to beginning of acoustic tokens.

- Have data sheet *Acoustic Tasks. Word Order ...* for this speaker nearby so you can see which of the tokens are acceptable and which are not.
- For numbering tokens, there should not be any gaps in the sequence. For example, if the 3rd token is not acceptable, number the 4th as 3 and the 5th as 4.

Result = 1, 2, 3, 4.

Not = 1, 2, 4, 5.

5. Start tape just before the speaker says the first token (as the examiner says the word in the background) and press *Enter*.
6. Stop digitizing (by pressing *Enter*) when 5–6 tokens have played or after just less than 20 sec. Try to record only complete tokens.
7. type **N** (for No).
8. Stop the tape.
  - you should now see a series of waveforms displayed on the screen.
9. Using the left and right mouse buttons, set the left and right cursors on either side of the token to be digitized.

AT ANY TIME, IF YOU WANT TO KNOW WHAT YOU'RE ISOLATING, JUST PRESS **P** AND THE COMPUTER WILL PLAY BACK TO YOU WHATEVER IS BETWEEN THE CURSORS (with no cursors present, it plays the whole screen).

- this may not work at all stages. It may be necessary to ack up to a previous screen in order to "play."
10. Press *Page Down*.
    - this will change the scale of the screen so that you only see the waveform of the token you just isolated.
  11. We don't need to store the entire phrase but we should include all of the /ēɪ/ of "say" (and a little of the /s/ noise) as well as the /ə/ in "again" (go to the burst for /g/).
    - use the cursors to isolate the target word as above.

FOR consonantal /r/ words, take the time to look at the spectrogram to determine if there is sufficient energy in F2 and F3 of /r/ to allow for determining the tracks of these formants. Do this before saving the token:

- f. press **A** (for Analysis).
  - g. press **T** (for Time/Freq).
  - h. enter **500** for BW (should only need to do this once).
  - i. press *Enter* (until Spectrogram appears).
  - j. press **S** (for Screen).
  - k. press **C** (for Colors).
    - if there is not enough energy present to tell where the tracks of F2 and F3 go during /r/, discard the token and renumber the other tokens as above.
  - l. To go back to the waveform only, press **M** (for Main Menu).
  - m. Repeat this process as you encounter each /r/ initial token.
12. Once you've isolated the word, you need to save it.
- type **F** ( for Files).
13. type **P** (for Put).
14. type **D** (for directory).
15. type **L:\acoustic\(speaker #)\word**
- skip steps 14 and 15 after the first token or each time you return to that file.
16. type **1**
17. type **(filename)** (see *Basic Filename Structure*).
- remember these should all end in *.wav*
  - press *Enter* (twice) to confirm.
18. Press *Page Up* to return to initial screen that shows the waveforms for all of the tokens.

19. Repeat steps 9–18 for next token.
20. Once all the tokens are digitized in that set, start again from step 2.
  - you should be able to omit step 4 after the first one.
21. Once all the tokens are digitized, proceed to next step or turn off CSpeech (see above).
22. Rewind tape.

### ***Measuring f0 and Word Durations***

The first analysis step involves measuring the speaker's fundamental frequency (f0) and the durations of all the tokens. Both of these may ultimately be used as covariates in the analysis with f0 being a general proxy for age and duration being a window on speaking rate.

1. Start CSpeech and get to main menu.
2. type **F** (for Files).
3. type **G** (for Get).
4. type **D** (for Directory).
  - only needed once at each logon or if you want to work on a different speaker's files.
  - on rest type **1** (and skip steps 5 & 7).
5. type **L:\acoustic\(speaker #)\word** (press *Enter*).
6. type **(filename)** (Press *Enter*).
7. type **1** (Press *Enter* to confirm).
  - a. f0 measurements.
    - these are taken on the tokens of the word "assign" only.
  - a1. locate the glottal pulsing for the /ēI/ in "say" on the waveform.
  - a2. using the left mouse button, place the left cursor at a point at least 5 glottal pulses to the left of /s/.

- a3. using the right mouse button, place the right cursor 50 msec (+/- 0.75 msec) from the left cursor.
  - a4. press *Page Down*.
  - a5. locate the 2 consecutive glottal pulses that look the most similar in form.
  - a6. place the left and right cursors as close as possible to the same spot on each of the two pulses.
  - a7. read  $f_0$  from top line of screen.
  - a8. repeat steps a1–a7 for the / $\bar{a}i$ / in "assign."
  - a9. repeat steps 2–7a8. for all tokens of "assign."
- b. Duration measurements: type **A** (for Analysis).
- b1. type **T** (for Time).
  - b2. Set BW (bandwidth) = **500**.
    - only need to set this for first item.
  - b3. Accept rest of settings by pressing *Enter* as needed.
    - after first item, you only have to press *Enter* once.
  - b4. type **S** (for Screen).
  - b5. type **C** (for Colors).
  - b6. Use left and right mouse buttons to isolate the precise start and end of the target word.
    - see APPENDIX A *Word Duration Criteria*.
  - b7. Write down the value shown in the top right corner of screen (Interval = \_\_\_\_\_) on the data sheet.
8. Repeat steps 2–7b7 for each word.
9. Proceed to next analysis step or turn off CSpeech (see above).

***Segmenting /s/***

These are extracted from the word files created above and stored in the **S** directory.

1. Start up CSpeech (see above) and turn on amplifier. Use the headphones and press **P** if you want to hear what you have isolated.
2. type **F** (for Files).
3. type **G** (for Get).
4. type **D** (for Directory).
5. type **L:\acoustic\(speaker #)\word**
  - after the first one, just type **word**
6. type **(filename)**.
7. type **1** (press *Enter* to confirm).
8. type **A** (for Analysis).
9. type **T** (for Time).
10. Enter BW = **500**.
11. Accept rest of setting by pressing *Enter* for each.
  - after 1st one, ignore step 10 & 11. Press *Enter* once.
12. type **S** (for Screen).
13. type **C** (for Colors).
14. Use left and right mouse buttons to isolate */s/*.
  - use APPENDIX B */s/ Segmentation Criteria*.
  - once isolated the segment needs to be saved.
15. type **M** (for Main Menu).
16. type **F** (for Files).

17. type **P** (for Put).
18. type **D** (for Directory).
19. type **s** (press *Enter*).
20. type **(filename)** (Press *Enter*).
  - this will be the same as the filename for the whole word and will be shown in green near the top of the screen.
21. type **1** (press *Enter*).
22. Repeat steps 2–21 for all tokens of all /s/ words.
23. Proceed to next step or turn off CSpeech (see above).

### ***Creating Moments Files***

Once all of the /s/ segments have been isolated and stored in the **S** directory, a detailed analysis known as "moments" must be carried out on them. This is done by CSpeech through a batch command.

1. type **F** (for Files).
2. type **D** (for DOS).
  - c:\csp will show up in a small sub-screen.
3. type **L:** (press *Enter*).
  - see which directory you are in. If it is the appropriate **S** directory for the subject you are working on:
  - type in the following batch command:
 

```
cbatch *.wav cmmt
```

CSpeech will now create the moments files (takes 1–2 minutes).
4. If you are not in the appropriate directory, get to that directory using the following:
  - type **cd\** (should now just read L:\).



- type **L:\acoustic\subject #)\s**
  - then enter the above batch command.
5. type **exit** (returns you to the CSpeech Main Menu).

### ***Analyzing Constriction Intervals***

This is the first step in the /r/ analysis. These are measured recorded and entered (later) into MINITAB by hand.

This first step is carried out on all /r/ words. It involves creating an LPC spectrum (which shows the location of the formants) and recording their center frequencies. The portion of the /r/ we are calling the constriction interval is the point where F2 and F3 are relatively flat. It is also sometimes called the "steady-state" portion.

1. Startup CSpeech, turn on amplifier and put on headphones.
2. type **F** (for Files).
3. type **G** (for Get).
4. type **D** (for Directory).
  - after the 1st one, type **1**
5. type **L:\acoustic\speaker #)\word**
  - ignore this step after 1st one for each speaker.
6. type **(filename)** (press *Enter*).
7. type **A** (for Analysis).
8. type **T** (for Time).
9. Set BW = **500**.
  - only need to *Enter* once.
10. Accept rest of settings by pressing *Enter* for each.
  - only press once after 1st one.
11. type **S** (for Screen).

12. type **C** (for Colors).
13. Locate the constriction interval of /r/ using the criteria below (in order of relative importance):
  - flat steady-state portion.
  - F2 and F3 come closer together.
  - use "play" to listen that it sounds /r/-like (i.e. not part of a vowel).
  - note also decreased intensity of higher formants relative to following vowel.

Isolate this portion using the left and right mouse buttons to control the cursors; in particular ***select the middle 20 msec*** of that section [ $\pm 0.75$  msec]. SEE ALSO NOTE AT END OF SECTION.

14. type **A** (for Analysis).
15. type **S** (for Spectrum).
16. type **L** (for LPC).
17. type **A** (for Analysis).
18. type **D** (for Display).

The mouse will now control a line that moves along the LPC spectrum displayed in the upper section of the screen.

19. Using the mouse, move the cursor along to the top of the second peak displayed. This is F2. While you are moving it note that a horizontal bar will be moving up and down on the spectrogram display below. Try to find the point where the horizontal bar is closest to the middle of the formant band and the cursor is closest to the top of the peak (the middle of the formant band being the more important of the two).
20. Record the Freq. value (shown to the left of the LPC spectrum display) on the data sheet.

Be aware that when you do a second LPC spectrum, two curves will appear. The green curve is always the most recent one and that will be the one you will make measurements from. The other curve is the 2nd most recent one and can be ignored for our present purposes.

21. Repeat 19–20 for F3 (3rd peak on the LPC spectrum).
22. Press *ESC*.

23. type **M** (for Main Menu).
24. Repeat steps 2–24 for all other /r/ tokens.
25. Proceed to next step or turn off CSpeech (see above).

**Note:** In some cases, there is no flat portion (the formants rise throughout the production). If that happens, measure both F2 and F3 at the point where F3 is the lowest (may not be the lowest point for F2 but measure F2 at that point anyway). To do this, set the cursors at 10 msec before and 10 msec after the lowest point and run LPC as above. To read the lowest point, set the horizontal cursor at the middle point of the lowest glottal pulse for that formant on the spectrogram (i.e. ignore the actual LPC waveform) and read the frequency from the LPC display as above.

### ***Creating Formant Track Files***

The transitions from /r/ to vowel and vowel to /r/ are an important part of /r/ analysis. In order to get this we need to track the pattern of the formants during the transitions. CSpeech has another batch command that estimates the path that the formants follow and stores these paths or "tracks" as digital files. To create these:

From the CSpeech Main Menu,

1. type **F** (for Files).
2. type **D** (for DOS).
  - the small sub-screen will again appear displaying: c:\csp
3. type **L:** (press *Enter*).
  - see which directory you are in. If it is the appropriate **WORD** directory for the subject you are working on:
  - type in the following batch command:
 

```
cbatch *.wav cfbw
```
  - this will again take several minutes.
4. If you are not in the appropriate directory, use the following to get to the directory you want:

type **cd\** (should now just read L:\).

type **cd\acoustic\(subject #)\word**

Then type the batch command (from Step 3).

5. type **exit** (press *Enter*).
  - this will return you to the CSpeech Main Menu.

### ***Correcting Formant Tracks (and measuring VOTs and burst frequencies)***

In creating the formant track files, CSpeech uses a mathematical algorithm that estimates where the center frequencies of the formants are throughout the production. As an estimate, it may have errors, so the tracks must be viewed and corrected. This is done on all /r/ tokens.

1. Startup CSpeech (see above).
  - then from Main Menu:
2. type **F** (for Files).
3. type **G** (for Get).
4. type **D** (for Directory).
  - only necessary for first token for one speaker. After that, type **1**
  - you can also skip step 5 after the first time.
5. type **L:\acoustic\(speaker #)\word**
6. type **(filename)** (press *Enter*).
7. type **1** (press *Enter*, to confirm).
  - skip after first time.
8. type **A** (for Analysis).
9. type **T** (for Time/freq).

10. Specify **BW = 500**.

11. Press *Enter* (after each of the other values).

- the first token, you can skip steps 10 and 11 by pressing *Enter* once

12. type **S** (for Screen).

13. type **C** (for Colors).

14. type **A** (for Analysis).

15. type **F** (for Formants).

16. type **E** (for Edit).

- this calls up the formant track files and overlays the estimated tracks onto the spectrogram as well as allowing you to edit the formant tracks.

Across the top of the screen (among other things) should be displayed F1, F2 and F3. One of them will be enclosed by arrows (e.g., >F2<). This is the active formant for editing. You can change this to edit any of the first three formants using the F1, F2 and F3 keys along the top of the keyboard.

**Note:** For the /r/ initial words, begin correcting each track at the beginning of the flat (stable) portion of /eI/ (this may be a different point for each formant). For the /r/ clusters, begin correcting the tracks at the point where voicing for the /r/ begins. For vocalic /r/, begin correcting the tracks at the start of voicing for the /b/. The end point for all tracks should be the next point where voicing stops (e.g., closure for a stop). For some words this may be the /ə/ in "again."

17. Begin with F1 by pressing the F1 key at the top of the keyboard.

18. Removing any formant tracks that extend beyond the limits of the tracks you want to correct. Isolate these using the left and right mouse buttons to set the left and right cursors and "zero" out as above.

Then type **Z** (for Zero).

- repeat for F2 & F3.

19. Beginning with F1, locate any areas where the formant track does NOT follow the center of the formant band. Isolate this section with the cursors (as above). This will "zero" out or delete that section of track:

Then type **Z** (for Zero).

20. If necessary, use the mouse to grab and drag the arrows at each end of the section and move it into the center of the formant band (use the left mouse button for the left arrow and the right mouse button for the right arrow).

**Caution:** Formant tracks are never "angular" (there should be no sudden direction changes), but rather they should be relatively smooth in their trajectories. In correcting tracks, you may need to put the track off center on occasion to smooth it out where direction changes occur.

21. type **I** (for Interpolate).
22. Repeat steps 17–20 for all areas where the track is not in the center of the formant band.

23. Press *Esc*

24. type **Y** (for Yes).

- this modifies the formant track files to include only the corrected values. These will be used to analyze transitions.

25. Repeat steps 19–24 for F2 and F3.

26. type **M** (for Main Menu).

27. For /r/ cluster tokens only, record Voice Onset Time (VOT), defined as the duration from the burst onset to the start of glottal pulsing (measured at F1). If there are pre-bursts, use the last burst (i.e. the one that is contiguous with the vowel).

- use the cursors to segment this interval and read the duration from the top line of the screen.

28. Also for /r/ cluster tokens only, record the burst onset frequency (i.e. the frequency locus of the burst). To do this isolate the segment from the burst (excluding pre-bursts) to 25 msec to the right of the burst with the cursors. Then:

type **A** (for Analysis).

type **S** (for Spectrum).

type **L** (for LPC).

type **A** (for Analysis).

type **D** (for Display).

A small horizontal cursor will appear on the spectrogram when you hold the right mouse button down and move it left and right. Find the most intense peak on the LPC display (highest peak present), move the cursor to the center of that peak using the right mouse button and read the frequency from the left side of the LPC display.

Note that the intensity values shown are in relative dB and thus are shown as negative values (the most intense element present in the file is given a value of 0 dB). Thus a peak of -35 dB is more intense than a peak of -40 dB.

Also note that if there are two peaks very similar to each other that might be the most intense (within 1.0 dB or less), record both.

29. Repeat steps 2–28 for the rest of the tokens.
30. Proceed to next step or turn off CSpeech (see above).

### ***Analyzing Formant Transitions***

This analysis is conducted outside of CSpeech in a Quattro Pro spreadsheet.

1. From the Windows 95 desktop, select Quattro Pro.
2. Select *File* from menu bar.
3. Select *acoustic\template.wb2* from current list.
  - if this is not in your current list:
    - select *open file*.
    - type **L:\acoustic\template.wb2**
    - put cursor into A2 cell.
4. Select *Notebook* from menu bar.
5. Select *text import*.
6. Select *drive*.
  - should be in **L:\acoustic** directory; if not:
    - select *L:\wcs\ptlab*.
    - select *acoustic*.
7. Select subject directory (e.g., *acoustic\c1\word*).

8. Enter filename (be sure it ends in *.fbw*).
  - you should now see data appear on the data page.
9. Put cursor into top cell (should show V>>>>V>>>>V>>>>...).
  - this line contains a set of commands that Quattro Pro uses to parse the data into appropriate columns.

**NEVER ERASE THIS LINE**

10. Hold down *Shift Key* and press *F7*.
11. Press *End*.
12. Press Down arrow.
13. Select *Notebook*.
14. Select *parse* (press *Enter*).
  - this tells Quattro Pro to organize the data into particular columns. Once organized, calculating formulas built into the spreadsheet template will carry out the calculations. The completed analysis then is available on the next parsed page.
15. Select parsed page from the tabs shown at the bottom of the page.
  - this page will be similar to the original page except there will be two additional columns, one for F2 and one for F3.
  - in the additional columns, you will find cells containing either '1' or '2' or nothing:
    - 1 = point where a rising transition was found (at 20 msec ahead, the freq. rose by at least 20 Hz).
    - 2 = point where a falling transition was found (at 20 msec ahead, the freq. fell by at least 20 Hz).

These are used as the basis for determining the extent and slope of the transitions. This analysis is only carried out on the /r/ initial tokens (i.e. singletons and clusters) and not on the vocalic /r/ tokens.

**Note:** TO BE A USABLE TRANSITION, IT MUST LAST AT LEAST 50 msec (anything less may be a product of random variation and/or the result of tracking errors). Thus any string of less than 10 consecutive 1s or 2s should be ignored.



Data to be recorded (for both F2 and F3):

<i>/r/ clusters</i>	offset transition (/r/ to vowel).
	offset delay (see notes).
<i>/r/ initial</i>	onset transition (vowel to /r/) + offset (/r/ to vowel).

For each transition:

16. Record the time and frequency of the first cell of a transition.
17. Record the time and frequency of the "cell after the last cell in the transition." This is the first point where the transition is no longer happening.
18. Once all the data has been recorded, erase all the data on the parsed page (except the columns containing 1s and 2s; where the formulas are located).
  - highlight columns A–E.
  - hold down *Shift* key and press *F7*.
  - press *Delete*.
19. Go to data page and erase data.
  - highlight A2 cell (the cell below V>>>>V>>>>...).
  - hold down *Shift* key and press *F7*.
  - press *Delete*.
20. Press the *Home* key.
21. Highlight cell A2.
22. Repeat steps 4–21 for next token.
23. Be sure to erase the data for the last token you do before exiting Quattro Pro.

—Notes —

1. On /r/ clusters only, if the offset transition starts more than 75 msec after the onset of voicing, please note this on the data sheet.
2. If you have any doubts about what is happening with the transitions, you can print out the parsed page and examine it while looking at the tracks on the spectrogram in CSpeech. To print a transition file:
  - highlight the portion to be printed.
  - select *Print*.
  - since the printout does not come with labels be sure to write the filename on it.
3. If you erase the parsing cell (V>>>>V>>>>...) by mistake, you can retrieve it by either:
  - select *Edit*.
  - select *UNDO Last*.
  - or *Exit* from Quattro Pro without saving anything.

### ***Entering Data into Database***

For specific procedures, please refer to separate document titled *Phonology Project Acoustic Database*.

### ***Reliability Measures*** (see also *Appendix C*)

To establish reliability levels, testing will be carried out on a sample of the control subjects. Both inter- and intra-judge reliability will be required. A sample of 3 control subjects will be randomly selected and a subset of tokens will be used to make a second set of measurements.

For each subject selected:

1. Re-calculate all 10 f0 values.
2. Randomly select 24 tokens.
  - re-measure durations.
3. Randomly select 10 /s/ tokens.
  - re-segment /s/.
  - create new moments files.
4. Generate a new set of formant track files.

5. Randomly select 12 /s/ tokens (singl., vocalic & cluster).
  - re-analyze constriction intervals.
  
6. Randomly select 10 /s/ tokens (exclude vocalics).
  - re-track formants.
  - re-analyze transitions.
  
7. Randomly select 10 /r/ cluster tokens.
  - re-measure VOTs.
  - re-measure burst frequencies.

## REFERENCES

Kent, R.D., & Read, C. 1992. *The Acoustic Analysis of Speech*. San Diego, CA: Singular Publishing Group, Inc.

Weismer, G. & Ingrisano, D. 1979. Phrase-level timing pattern in English: Effects of emphatic-stress location and speaking rate. *Journal of Speech and Hearing Research*, 22, 3, 516–533.

**APPENDIX A**  
***Word Duration Criteria***

(see also notes to follow)

Please note that while some of these may not correspond to "words" in the true sense, they are based on historical precedents in the acoustics literature (cf. Weismer & Ingrisano, 1979).

<i>assign</i>	from 1st glottal pulse of /ə/ to last glottal pulse associated with /n/ (where F2 is still visible).
<i>bird, burg, burr</i>	from the burst release associated with /b/ to the last glottal pulse for /r/ (the onset of closure for the stop, if applicable).
<i>cosine</i>	from burst release for /k/ to last glottal pulse associated with /n/.
<i>cried, crude, pried, prude, tried</i>	from the burst release of the stop to the last glottal pulse of the vowel/diphthong.
<i>kicks, kiss, kits</i>	from the burst release of /k/ to the first glottal pulse of the /ə/ in "again." • if noise is not continuous with the /ə/, word ends at last of frication, using high freq. noise as /s/ marker.
<i>kin, pin</i>	from the burst release for /p/ or /k/ to the last glottal pulse associated with /n/.
<i>rebel (both versions)</i>	from 1st glottal pulse of /r/ to the last glottal pulse of /l/.
<i>ride, rude</i>	from the 1st glottal pulse of /r/ to the last glottal pulse of the vowel or diphthong (the onset of closure for the stop).
<i>sin, skin, soon, spin, spoon</i>	from the last glottal pulse of /eɪ/ in "say" to the last glottal pulse of /n/. • if preceding vowel is not continuous with /s/, start from the onset of frication, using high freq. noise as /s/ marker.

**Note 1. Defining /r/.** The problem here is deciding when /ēī/ in "say" ends and /r/ begins. Use the following clues (in order of relative importance):

- point where F2/F3 transitions flatten out.
- acoustic feedback; play what you have and listen to be sure there is no vowel present.
- watch for drop in intensity of F1.

**WHEN IN DOUBT, ALWAYS MOVE AWAY FROM THE VOWEL.**

**Note 2.** For words ending in voiced sounds (especially sonorants e.g., /l,n,r/), there may be no clear break between that final segment and the word "again." Here use the last glottal pulse prior to the rise in F2 of the /ə/ in "again." Likewise when there is no break between "say" and "assign," use the last glottal pulse prior to the F2 fall in the /ə/ of "assign."

**Note 3.** Start of closure for stops should be the point of "true closure." There should be no resonances present in the interval. The exception could be some glottal pulsing which of course might be seen with voiced stops.

**Note 4.** Occasionally productions of stops may include "pre-bursts." These look like the release of the stop but do not release into the vowel. These should be ignored. The beginning of a stop for our purposes will be defined as the burst that CLEARLY RELEASES INTO THE FOLLOWING VOWEL. Use waveform (oscillogram) to assist if needed; a spike will usually correspond to true burst.

**APPENDIX B**  
***/s/ Segmentation Criteria***

<i>Assign</i>	from the last glottal pulse of /ə/ or the beginning of frication* to the first glottal pulse of /ɑ̃/ or the end of frication.*
<i>Cosine</i>	from the last glottal pulse of /ou/ or the beginning of frication* to the first glottal pulse of /ɑ̃/ or the end of frication.*
<i>Kicks, kits</i>	from the burst release of /t/ (or /k/) to the end of frication* or the first glottal pulse of /ə/ in "again."
<i>Kiss</i>	from the last glottal pulse of /ɪ/ or the beginning of frication* to the first glottal pulse of /ə/ in "again" or the end of frication.*
<i>Sin, soon</i>	from the last glottal pulse of /ẽ/ in "say" or the beginning of frication* to the first glottal pulse of following vowel or the end of frication.*
<i>Skin, spin, spoon</i>	from the last glottal pulse of /ẽ/ in "say" or the beginning of frication* to the onset of closure of the stop following /s/.

\* *glottal pulse criteria preferred unless frication is not continuous with the vowel*

## **APPENDIX C**

### ***Measurement Precision***

The reader should refer to Kent and Read (1992, pp. 61–66) for a more detailed discussion of this issue, but the following is intended to highlight the problem. CSpeech (like other digital analysis software) reports data with what appear to be incredible levels of precision (e.g., up to three decimal places). The precision of the data should, however, be approached with some caution. First, the ability of the user to accurately position the cursor at the precise point on a display is subject to some measurement error. Second, the mouse itself has physical constraints as to how finely it is able to respond to the user's movements. And third, precision is compounded at every step as a function of the size of the analysis window. For example, CSpeech allows the user to isolate a portion of an utterance and then *Page Down* to expand the view of that portion so that it is displayed on the entire screen. And this process can be repeated multiple times. Thus, for the entire utterance the screen may represent 250 msec while for the expanded portion of the utterance it may represent only 2.5 msec. If the mouse can only provide resolution of say 1/100 of the screen width, this represents a precision of +/- 2.5 msec for the whole utterance, but +/- .025 msec for the expanded portion (a difference of 2 orders of magnitude). Any measurements made using these tools must carefully account for all of these sources of variance before decisions are made as to how reliable the values obtained will be and how data should be reported.