
RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 20 (1995)
Indiana University

**Acoustic and Glottal Excitation Analyses of
Sober vs. Intoxicated Speech: A First Report¹**

Kathleen E. Cummings,² Steven B. Chin, and David B. Pisoni

*Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405*

¹This work supported in part by grants to Indiana University from the Alcoholic Beverage Medical Research Foundation and the National Institutes of Health (NIH-NIDCD), Training Grant DC00012.

²Also Digital Signal Processing Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA.

Acoustic and Glottal Excitation Analyses of Sober vs. Intoxicated Speech: A First Report

Abstract: This is a first report of results from acoustic and glottal excitation analyses of speech produced both with and without alcohol. The new analyses were designed to determine whether there are significant and identifiable differences in phonation between the two types of speech. Non-nasal vowels extracted from eight isolated words produced by four talkers in both a nonalcohol and an alcohol condition were examined in terms of (1) direct measures of acoustic speech waveform parameters, (2) perturbation measures of acoustic speech waveform parameters, and (3) measures of the glottal excitation waveshape. Parameters related to the steadiness of speech production, as reflected in perturbations in adjacent pitch periods, exhibited differences between alcohol and nonalcohol speech. Specifically, we found consistent differences between the two types of speech on several measures of jitter, although the amount of variation between the alcohol and nonalcohol speech appeared to be talker-dependent.

Introduction

The goal of the research reported here was to analyze sober versus intoxicated speech in order to determine whether speech produced while a person is intoxicated is significantly and identifiably different from speech produced while a person is sober. To this end, several measures related to voiced excitation have been extracted and studied for four of the speakers in the Indiana University Alcohol Speech Database (Pisoni & Martin, 1989; Pisoni, Yuchtman, & Hathaway, 1986). These measures can be divided into three categories: direct measures of the acoustic waveform, perturbation measures of the speech waveform, and measures of the glottal excitation waveform. Because of physiological differences between speakers, most of these parameters may vary a great deal from speaker to speaker. The present research effort attempts to identify significant parameter variation trends in intoxicated versus sober speech for a particular speaker that are consistent for all of the speakers studied.

Of all of the parameters we have extracted thus far, those that are related to the steadiness with which a person produces speech are the parameters that best reflect the differences between intoxicated and sober speech. For example, several measures of jitter appear to be consistently different for intoxicated versus sober speech. The amount of variation between sober and intoxicated speech also seems to be speaker-dependent. In one speaker in particular, KM, all parameters vary less drastically between sober and intoxicated speech. We are currently investigating the possibility that he is a tolerant drinker.

Method

Materials for the acoustic and glottal excitation measurements described here were taken from a digital database of isolated monosyllabic words, isolated spondaic words, isolated sentences, and connected passages spoken by nine young adult male talkers in two conditions: without alcohol and under alcohol at .10% BAC or higher. Full details regarding talker selection and preparation, speech materials, and

elicitation procedures can be found in Pisoni et al. (1986) and Pisoni and Martin (1989). Analyses of up to four talkers are reported here. Table 1 shows subject data for these four talkers, including age, initial BAC, final BAC, and self-reported alcohol intake.

Table 1

Talker Characteristics

Age = age in years at last birthday; Initial BAC = BACs at beginning of recording in g/100 ml blood, as measured by Smith & Wesson Breathalyzer (Model 900A); Final BAC = BACs at end of recording; Alcohol Intake = self-reported total alcohol intake during 30 days prior to recording session, converted to oz 200-proof alcohol. (From Pisoni, Yuchtman, & Hathaway, 1986.)

Talker	Age	Initial BAC	Final BAC	Alcohol Intake
DP	21	0.15	0.10	8.94
JB	26	0.10	0.10	6.15
JS	22	0.16	0.10	3.53
KM	21	0.17	0.10	16.80

Speech tokens of isolated words were elicited in a shadowing task in both the alcohol and nonalcohol conditions; auditory stimuli were presented via audio tape playback over headphones, and talkers were instructed to simply repeat words aloud as quickly as possible. Vowels from eight isolated words ('chaff' (Word 11), 'chap' (Word 12), 'cheese' (Word 13), 'chest' (Word 14), 'chief' (Word 15), 'choose' (Word 17), 'chops' (Word 18), and 'heath' (Word 61)) in each of the two conditions, sober and intoxicated, were used for the analyses described here. All of the parameters have been extracted and studied for four speakers designated DP, JB, JS, and KM.

Each utterance was pitch-marked using a semi-automatic cepstrum-based pitch detector on a Sun 4 or Sun SPARCstation 20. The boundaries of the voiced sections were marked, and a pitch contour was determined. Parameters were then extracted using each original utterance, the voicing boundaries, and the pitch contours. Three types of measures of voiced excitation were extracted:

1. direct measures of acoustic speech waveform parameters,
2. perturbation measures of acoustic speech waveform parameters, and
3. measures of the glottal excitation waveshape.

The pitch contours (in samples, sampling rate of 20 kHz) are shown in Figures 1-8. Each figure contains four plots, each of these comparing the pitch contours for sober and intoxicated versions of the same word for a given speaker.

Insert Figures 1 through 8 about here

Parameters

Direct Measures of the Acoustic Speech Waveform

Typically, parameters that are used to distinguish between different speaking styles involve measures of the energy, or RMS intensity, in a given segment of speech, and measures of the pitch period (see Cummings, 1992). Several such parameters were extracted directly from the acoustic speech waveforms. These included:

1. **pave**: mean of the pitch contour for an utterance
2. **pc**: measure of the flatness of the pitch contour
3. **aint**: mean(total RMS intensity per pitch period) for an utterance
4. **mint**: max(total RMS intensity per pitch period) for an utterance
5. **naint**: mean((total RMS intensity per pitch period)/(pave)) for an utterance
6. **nmint**: max((total RMS intensity per pitch period)/(pitch at the max location)) for an utterance.

Sample distributions were determined for each of these parameters for each speaker in each of the two speaking conditions (sober and intoxicated).

Summary of Results. Results from direct measures of the acoustic waveform for the four talkers were as follows:

- The average pitch was higher for intoxicated speech than for sober speech. Average pitch was measured in samples; thus, the frequency of the average pitch was lower for intoxicated speech than for sober speech.
- The average pitch was more tightly clustered (i.e., varied less about the mean) for sober speech than for intoxicated speech.
- The pitch contour measure was somewhat mixed: more flat in sober speech for three speakers (DP, JB, and JS) but more flat in intoxicated speech for one speaker (KM). This result can also be observed by looking at the pitch contours in Figures 1–8. As a rule, the intoxicated and sober pitch contours had similar general shapes for a given word.
- As expected, all four of the RMS intensity measures exhibited the same behavior. Like the pitch contour measure, the results were mixed: RMS intensity was lower in sober speech for three speakers (DP, JB, and KM) but higher for the fourth speaker (JS). For speaker JS, word 15 ('chief') in the sober condition had a much higher RMS intensity than any other word.

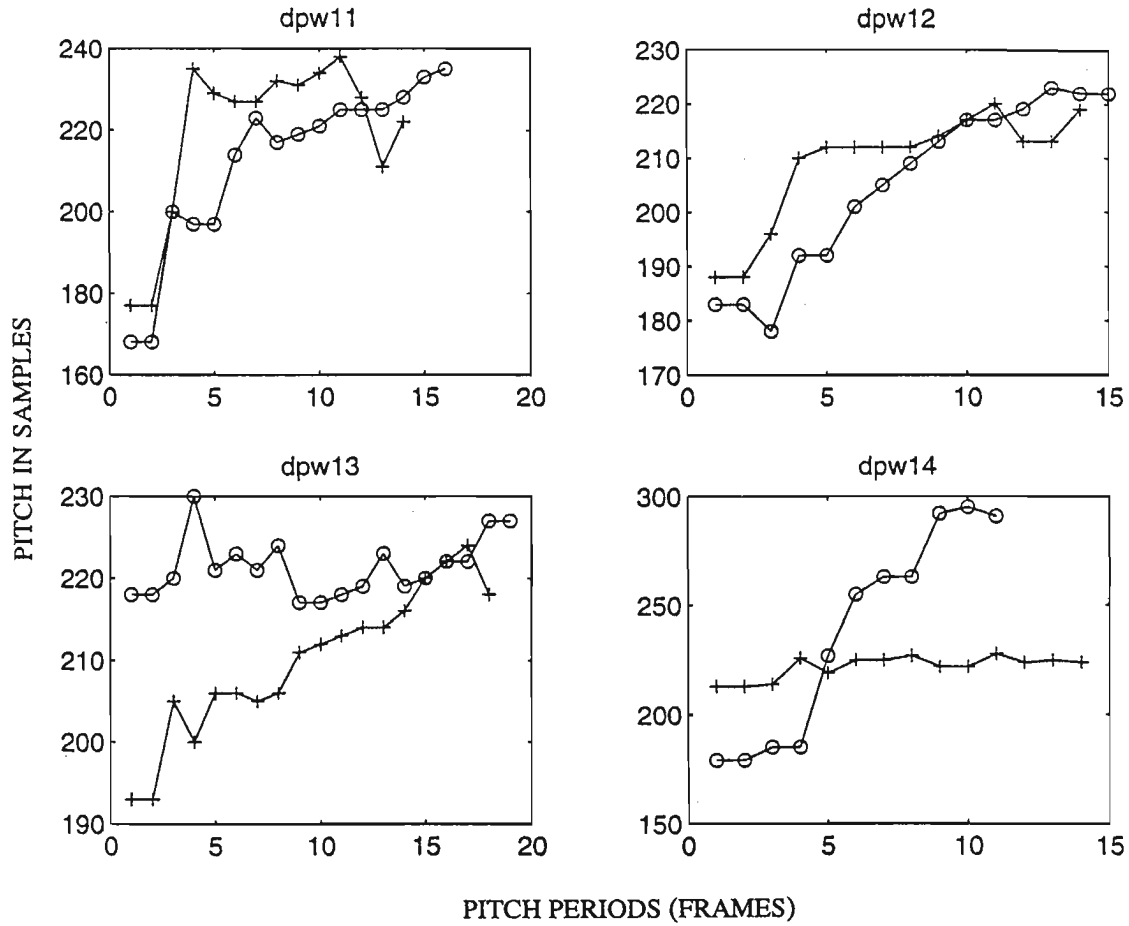


Figure 1: Sober (+) and intoxicated (o) pitch contours for subject DP for the words 'chaff' (11), 'chap' (12), 'cheese' (13), and 'chest' (14).

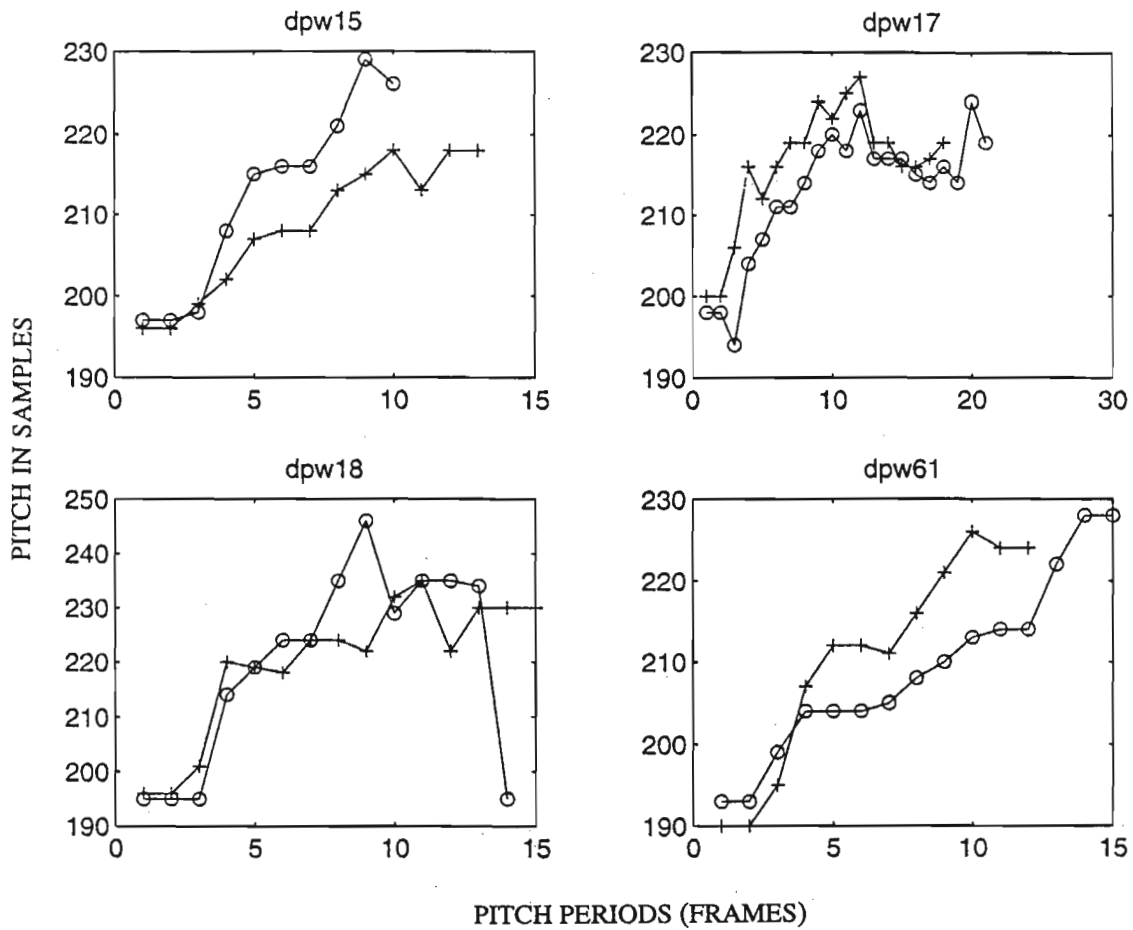


Figure 2: Sober (+) and intoxicated (o) pitch contours for subject DP for the words 'chief' (15), 'choose' (17), 'chops' (18), and 'heath' (61).

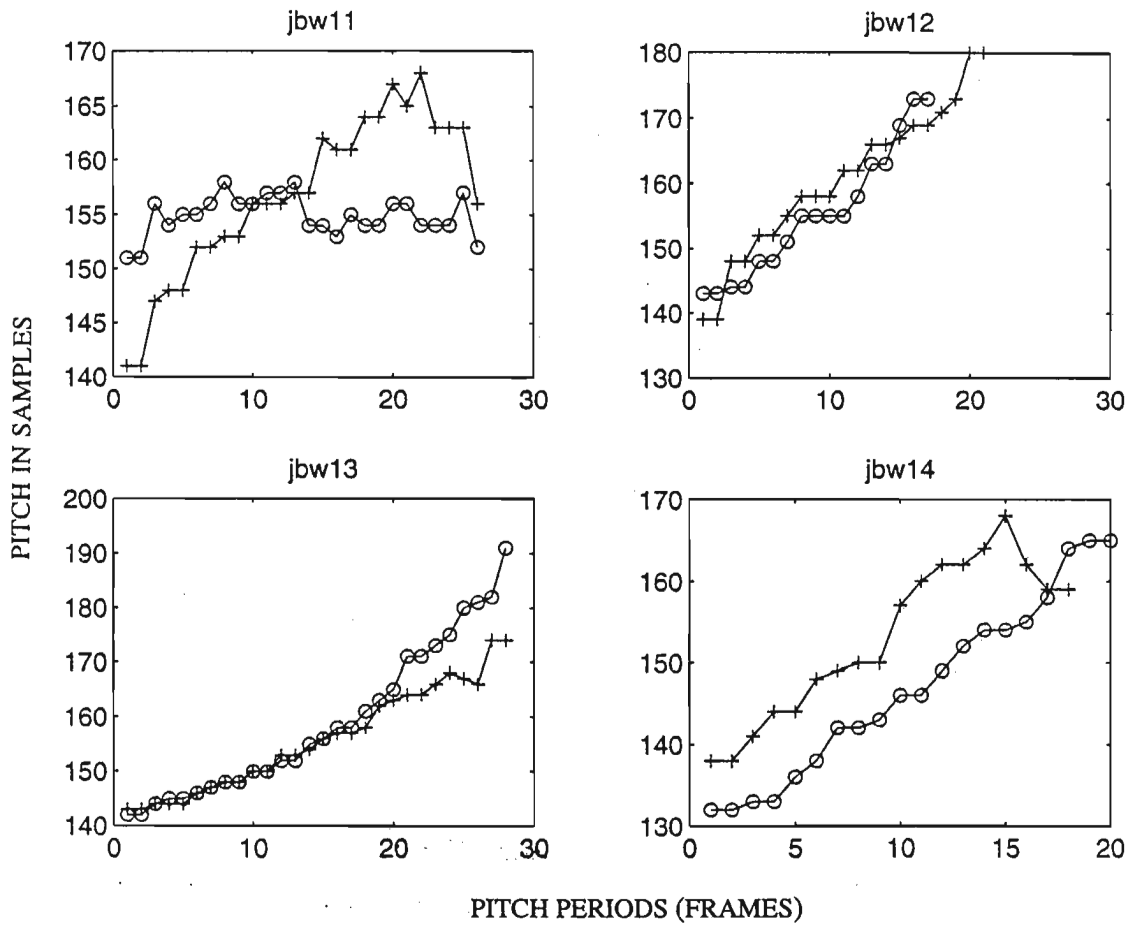


Figure 3: Sober (+) and intoxicated (o) pitch contours for subject JB for the words 'chaff' (11), 'chap' (12), 'cheese' (13), and 'chest' (14).

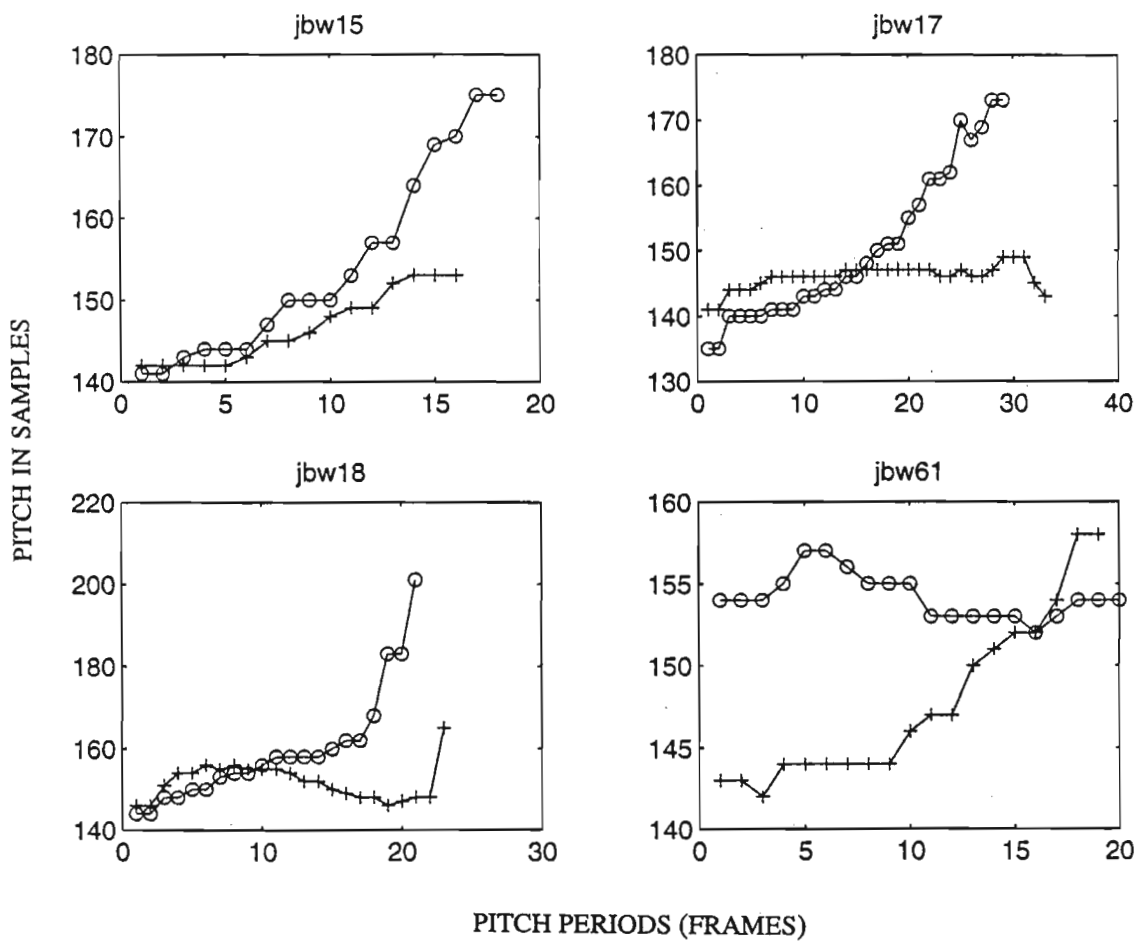


Figure 4: Sober (+) and intoxicated (o) pitch contours for subject JB for the words 'chief' (15), 'choose' (17), 'chops' (18), and 'heath' (61).

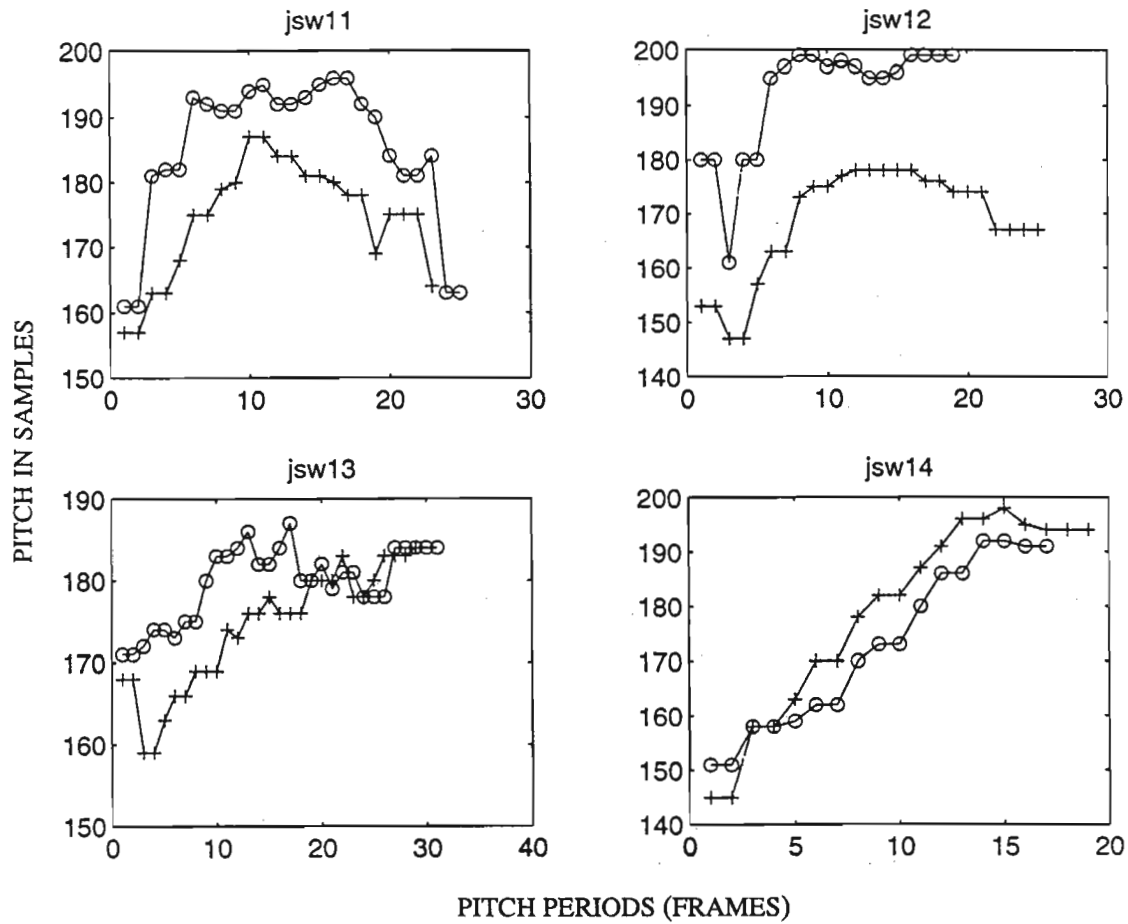


Figure 5: Sober (+) and intoxicated (o) pitch contours for subject JS for the words 'chaff' (11), 'chap' (12), 'cheese' (13), and 'chest' (14).

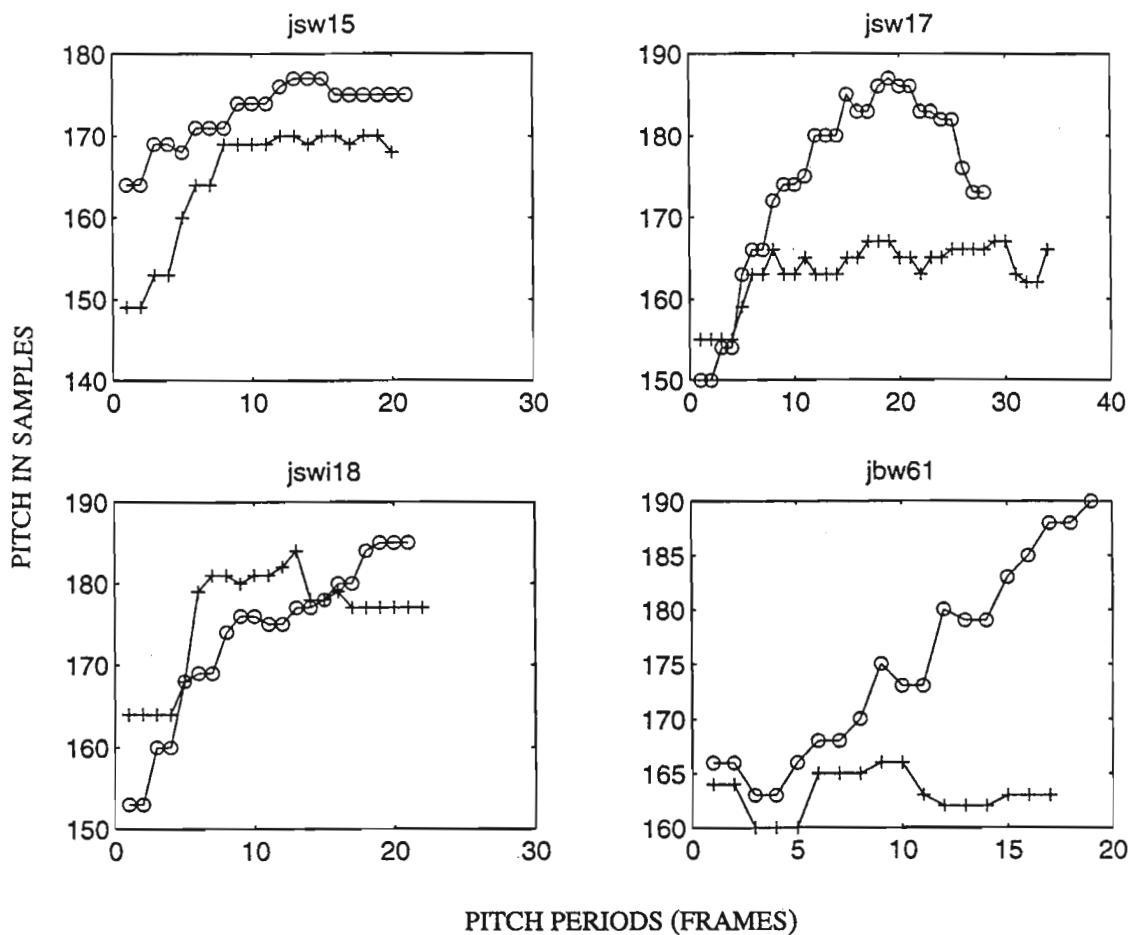


Figure 6: Sober (+) and intoxicated (o) pitch contours for subject JS for the words 'chief' (15), 'choose' (17), 'chops' (18), and 'heath' (61).

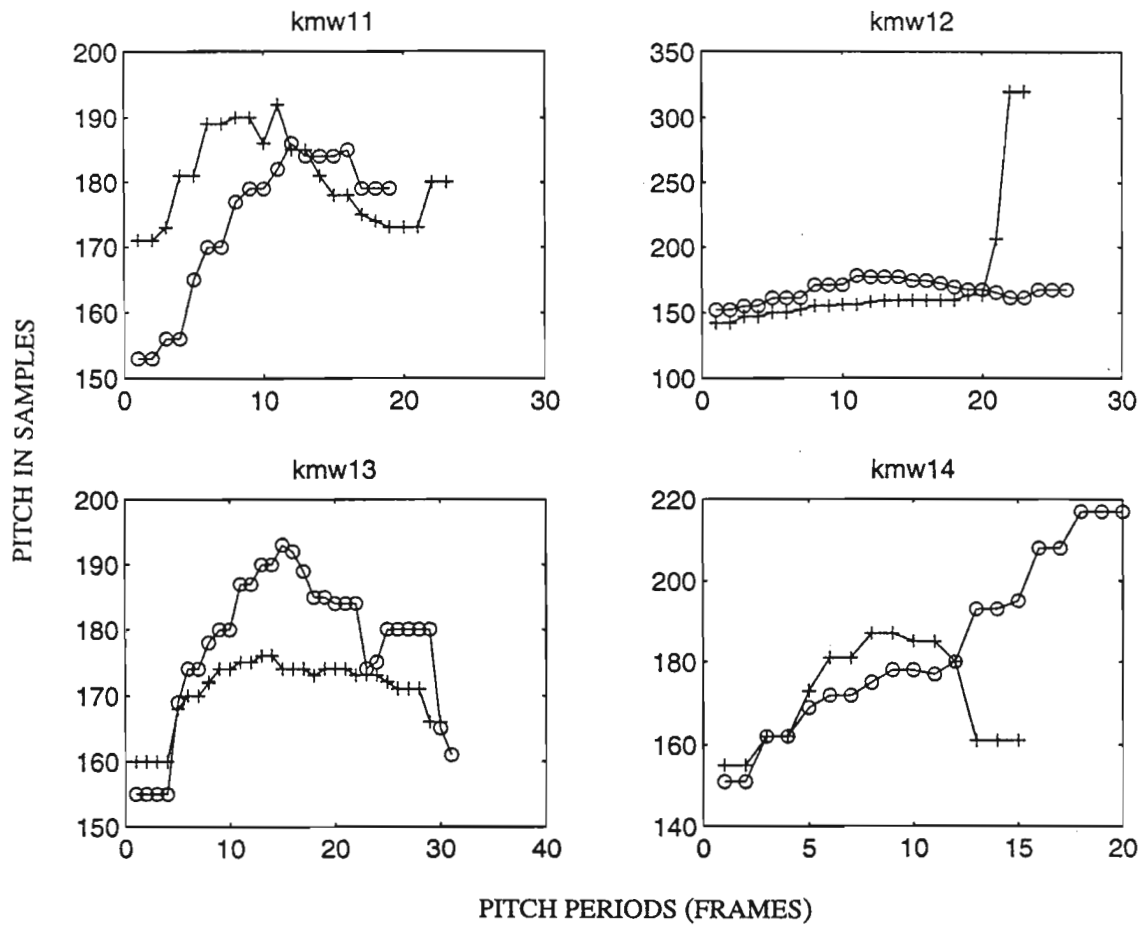


Figure 7: Sober (+) and intoxicated (o) pitch contours for subject KM for the words 'chaff' (11), 'chap' (12), 'cheese' (13), and 'chest' (14).

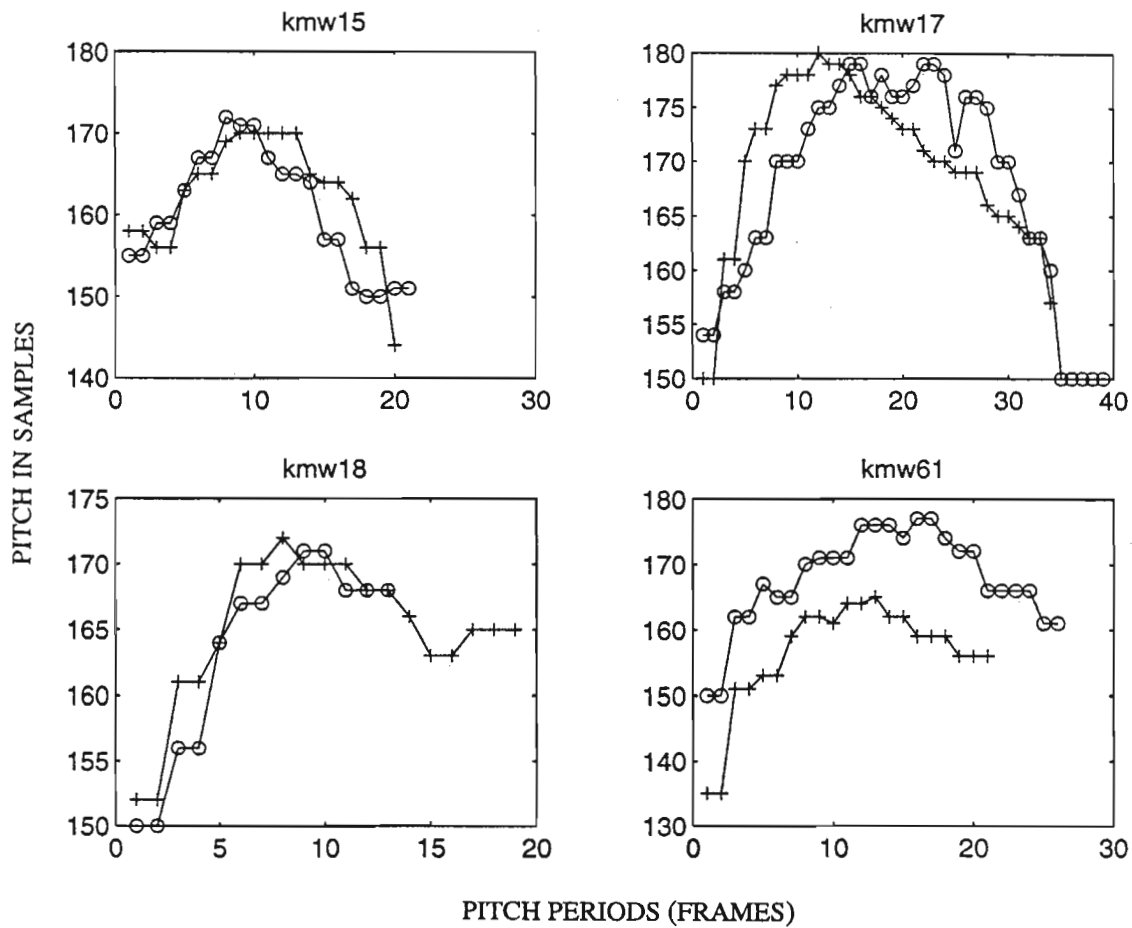


Figure 8: Sober (+) and intoxicated (o) pitch contours for subject KM for the words 'chief' (15), 'choose' (17), 'chops' (18), and 'heath' (61).

Perturbation Measures of the Acoustic Speech Waveform

In addition to direct measures of speech production such as those discussed in the preceding section, a number of perturbation measures were also extracted and examined in detail. Perturbation measures reflect the steadiness with which one produces speech. Since alcohol use affects a person's motor control (see Starmer, 1989), it is reasonable to expect perturbation measures to reflect differences in the speech production process when a person is intoxicated as compared to speech produced in the sober condition.

Initially, a series of perturbation parameters based on the six direct measures of the acoustic speech waveform discussed above were extracted. These were, generally, measures of the magnitude and direction of change in pitch or RMS intensity from pitch-period to pitch-period. The resulting parameter distributions displayed increases in variability in the intoxicated over the sober speech conditions.

A full perturbation analysis used measures based on those suggested by Pinto and Titze (1990). There are many definitions of 'jitter' and 'shimmer' in the literature (see Baken, 1987), but in a general sense, jitter is a measure of pitch frequency variability. In other words, jitter measures how much the pitch frequency changes from pitch-period to pitch-period. Shimmer is a measure of the change in energy from pitch-period to pitch-period.

In order to measure jitter and shimmer, in fact, in order to measure the changes in pitch and energy in a variety of ways, perturbation analysis was performed on three parameters:

$$F0_{val} = \frac{F0}{\text{mean}(F0) \text{ over the utterance}}$$

$$Amp_{val} = \frac{|\text{mean}(\text{amplitude in a pitch period})|}{\text{mean}(|\text{mean}(\text{amplitude in a pitch period})|) \text{ over the utterance}}$$

$$Int_{val} = \frac{\sqrt{\sum_{i=1}^L s^2(i)}}{\text{mean}(\sqrt{\sum_{i=1}^L s^2(i)}) \text{ over the utterance}}$$

where L is the length of a given pitch period.

Perturbation measures of $F0_{val}$ are related to jitter, while perturbation measures of Amp_{val} and Int_{val} are related to measures of shimmer.

Standard perturbation analysis was carried out on each of these three parameters. Zero-th, first, and second order perturbation vectors were calculated for each of the three parameters for each utterance. Letting a_i be the cyclic parameter (either $F0_{val}$, Amp_{val} , or Int_{val}) in the i th cycle of N total cycles of the waveform

$$\bar{a} = \frac{1}{N} \sum_{i=1}^N a_i$$

$$p_i^0 = a_i - \bar{a} \quad \text{for } i=1, \dots, N$$

$$p_i^1 = p_i^0 - p_{i-1}^0 = a_i - a_{i-1} \quad \text{for } i=2, \dots, N$$

$$p_i^2 = p_{i+1}^1 - p_i^1 = a_{i+1} - 2a_i + a_{i-1} \quad \text{for } i=2, \dots, N-1$$

For each perturbation vector (p^0 , p^1 , and p^2) for each of the three parameters (F0val, Ampval, and Intval) for each utterance, four perturbation measures were calculated. N_k is the length of the k th order perturbation vector, p^k , and $k = 0, 1$, and 2 . These four measures are

1. MR_k – rectified mean (centroid of the histogram)
2. MER_k – median (not rectified)
3. rms_k – root-mean-squared value, where

$$rms_k = \sqrt{\frac{1}{N_k} \sum_{i=1}^{N_k} p^k(i) \cdot p^k(i)}$$

4. ZCR_k – zero-crossing rate (number of sign changes in p^k divided by $N_k - 1$)

The first three values are measures of perturbation extent; the last value, the zero-crossing rate, is a measure of perturbation rate.

A number of common measures of jitter and shimmer can be related to these four measures of the perturbations when the original parameter is pitch frequency or energy, respectively. For example, measures of jitter from Hollien, Michel, and Doherty (1973) and Jacob (1968) can be related to MR^1/\bar{a} , while Ludlow, Coulter, and Gentges's (1983) deviation from linear trend and Koike's (1973) relative average perturbation can be related to MR^2/\bar{a} . It is thought that the ratios rms^1/rms^0 , rms^2/rms^0 , and rms^2/rms^1 are related to the temporal nature of the perturbations.

Results of the perturbation analysis are shown graphically in Figures 9 through 17. Each graph compares the sample distributions for each speaker for intoxicated (I) versus sober (S) speech. The mean and plus and minus one standard deviation are shown. These include the following:

1. $rms_0 - F0val$
2. $MR_0 - F0val$
3. $MER_1 - F0val$
4. $MER_0 - F0val$
5. $ZCR_0 - Ampval$
6. coefficient of variation - F0val
7. direct measure of shimmer - Ampval
8. period variability index - F0val
9. $rms_1/rms_0 - F0val$

 Insert Figures 9 through 17 about here

Summary of Results. The results involving F0val were larger than the results involving either Ampval or Intval. A number of the perturbation measures involving F0val showed the same variation from sober to intoxicated speech for all four speakers. For example, rms_0 for F0val, which is a measure of jitter rate, was consistently higher in intoxicated speech than in sober speech for all four speakers. As another example, MR_0 for F0val, which is a measure of jitter extent, was also higher in intoxicated speech than in sober speech for all four speakers. The coefficient of variation and the period variability index, two other measures of the extent of jitter, were also higher in intoxicated speech than in sober speech for all four speakers. Also, we found that one speaker, KM, showed less change in intoxicated versus sober speech for all measures.

Interestingly, the one exception to this rule was that the rms-ratio parameters (which are believed to be related to the temporal nature of the perturbations) were significantly different only for speaker KM. We are currently investigating the possibility that this speaker was a more tolerant drinker. Many of the results were not consistent across all four speakers but did show significant differences between sober and intoxicated speech.

Glottal Excitation Waveshape

The final portion of this phase of the project involved analyzing the glottal excitation waveshapes of sober versus intoxicated speech. Thus far, glottal waveforms have been extracted from two speakers (DP and JB) for two utterances. The speech was downsampled to 10 kHz prior to inverse-filtering. The glottal waveforms were then extracted using an adaptation of Wong, Markel, and Gray's (1979) closed phase glottal analysis. In this method, glottal closure is roughly identified from the covariance linear prediction

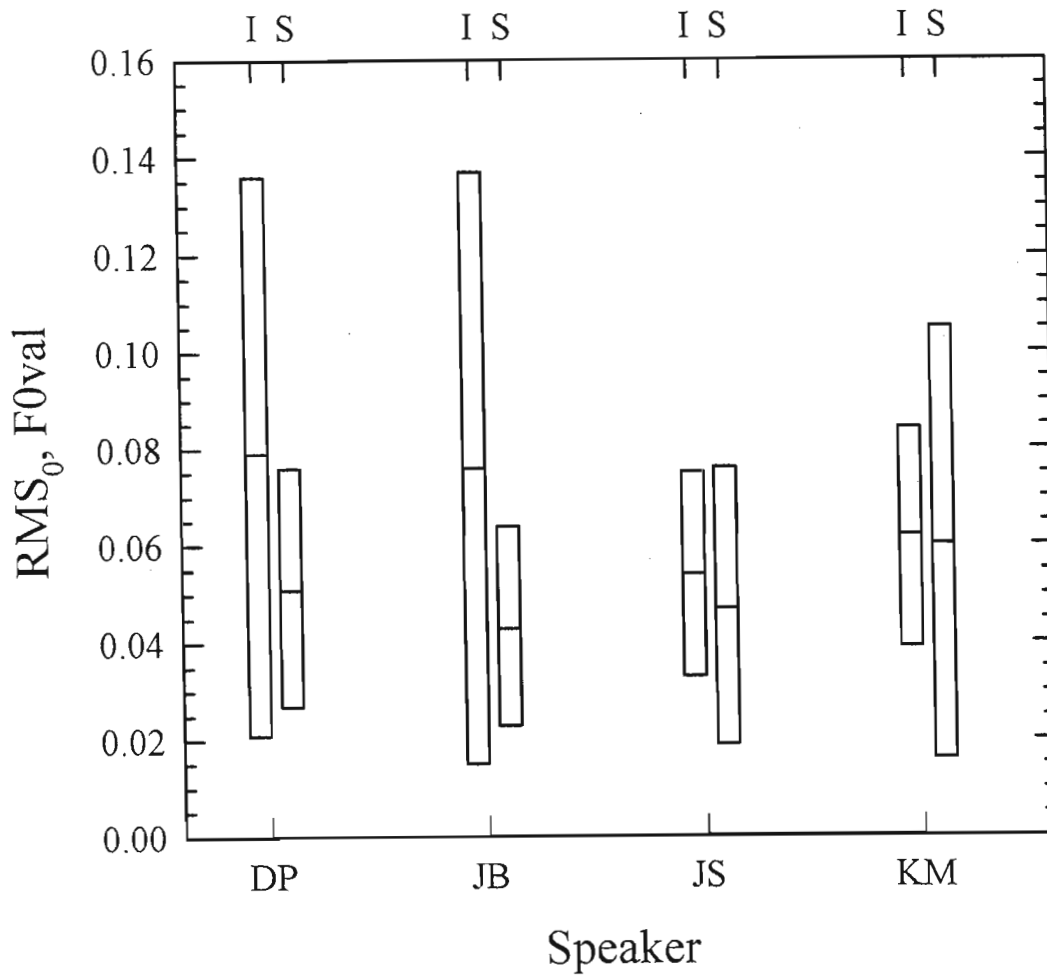


Figure 9: Sample distributions (mean and \pm one standard deviation) for 4 speakers in intoxicated (I) and sober (S) conditions for perturbation measure $RMS_0, F0val$.

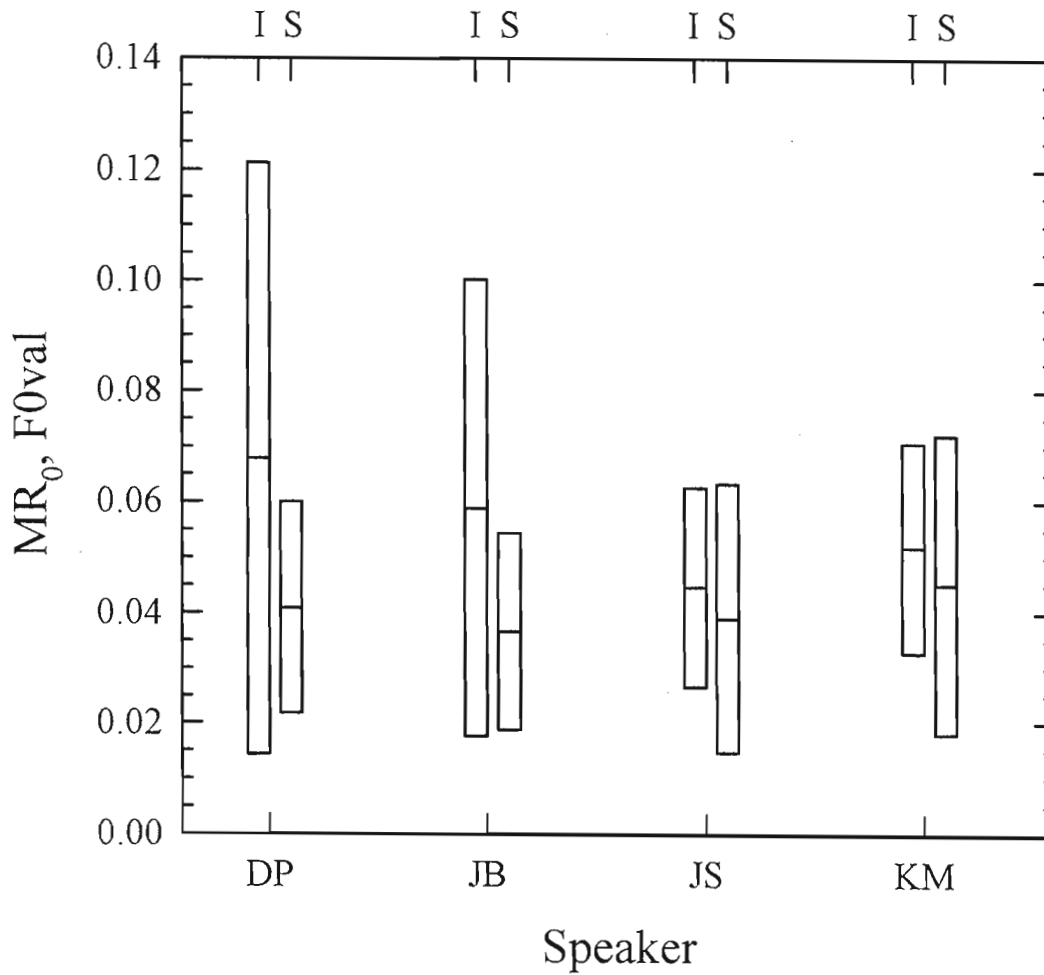


Figure 10: Sample distributions (mean and \pm one standard deviation) for 4 speakers in intoxicated (I) and sober (S) conditions for perturbation measure $MR_0, F0val$.

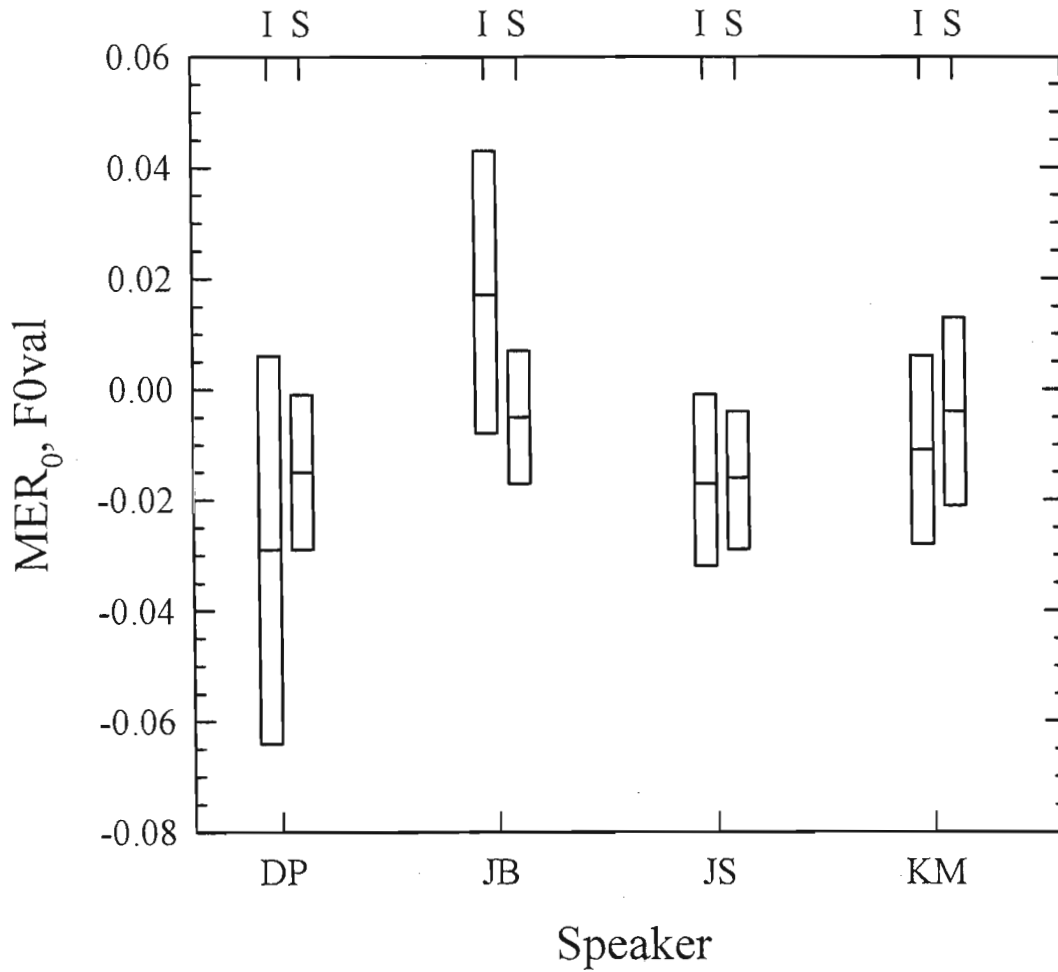


Figure 11: Sample distributions (mean and \pm one standard deviation) for 4 speakers in intoxicated (I) and sober (S) conditions for perturbation measure $MER_0, F0val$.

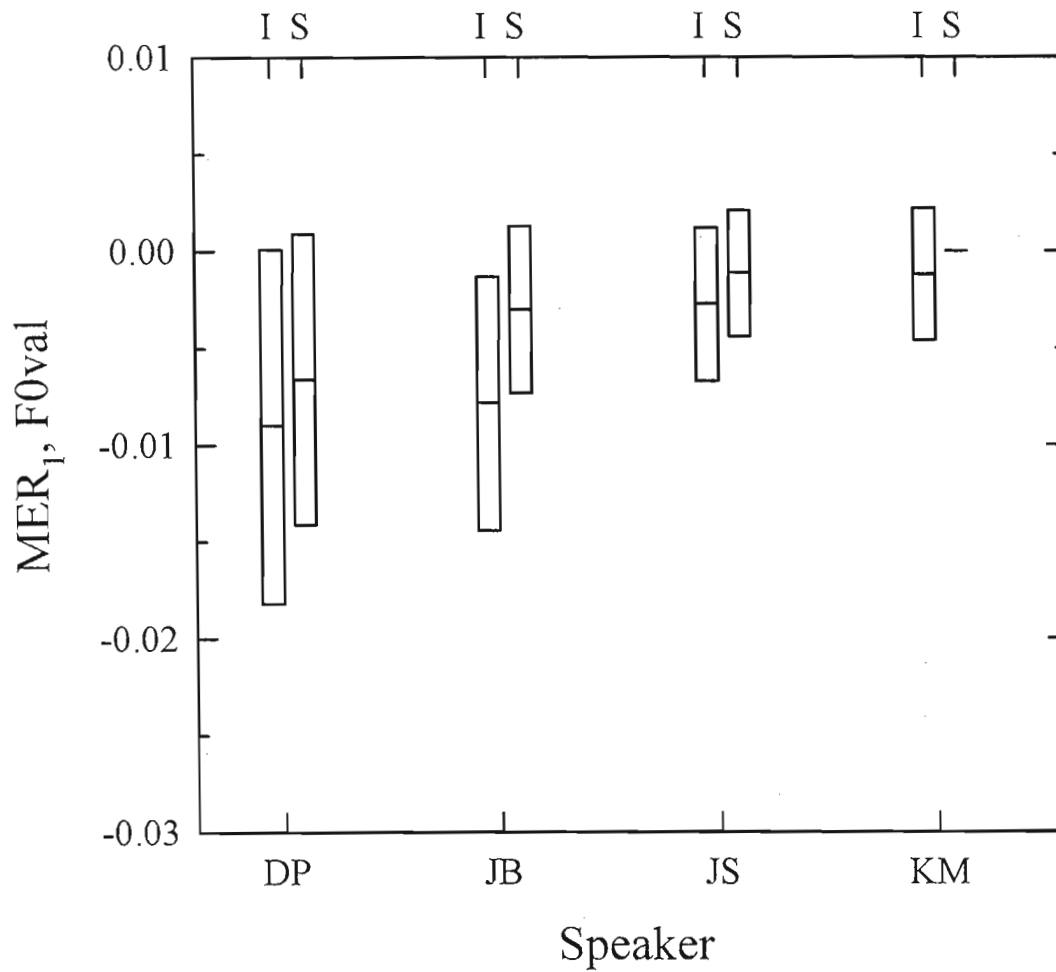


Figure 12: Sample distributions (mean and \pm one standard deviation) for 4 speakers in intoxicated (I) and sober (S) conditions for perturbation measure $MER_1, F0val$.

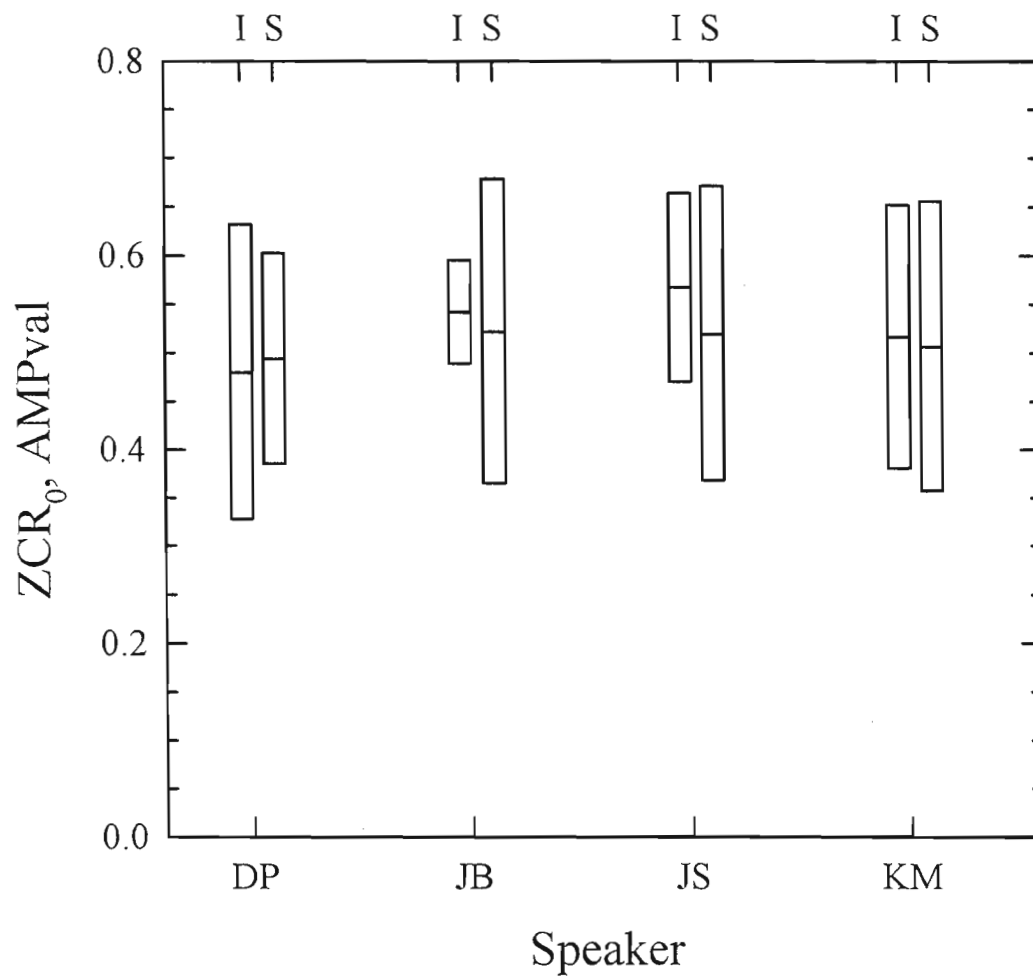


Figure 13: Sample distributions (mean and \pm one standard deviation) for 4 speakers in intoxicated (I) and sober (S) conditions for perturbation measure ZCR_0 , AMPval.

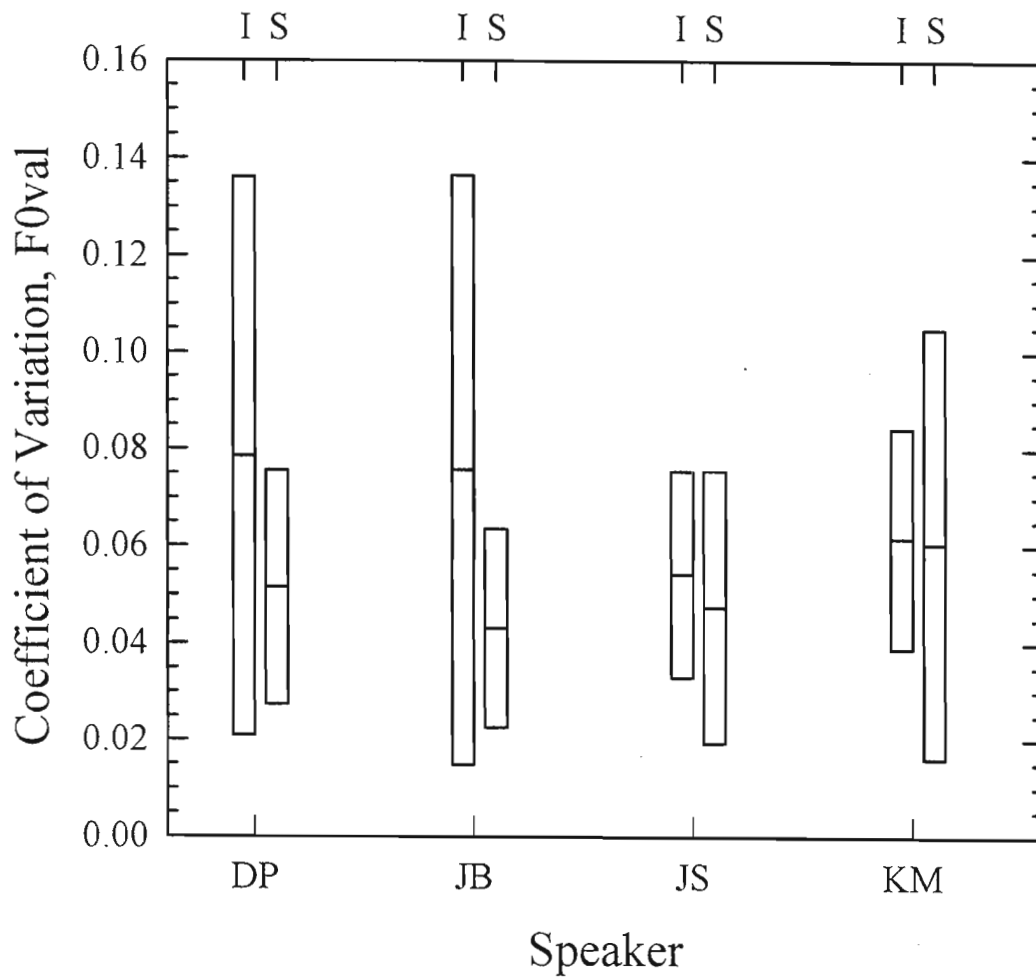


Figure 14: Sample distributions (mean and \pm one standard deviation) for 4 speakers in intoxicated (I) and sober (S) conditions for perturbation measure Coefficient of Variation, F0val.

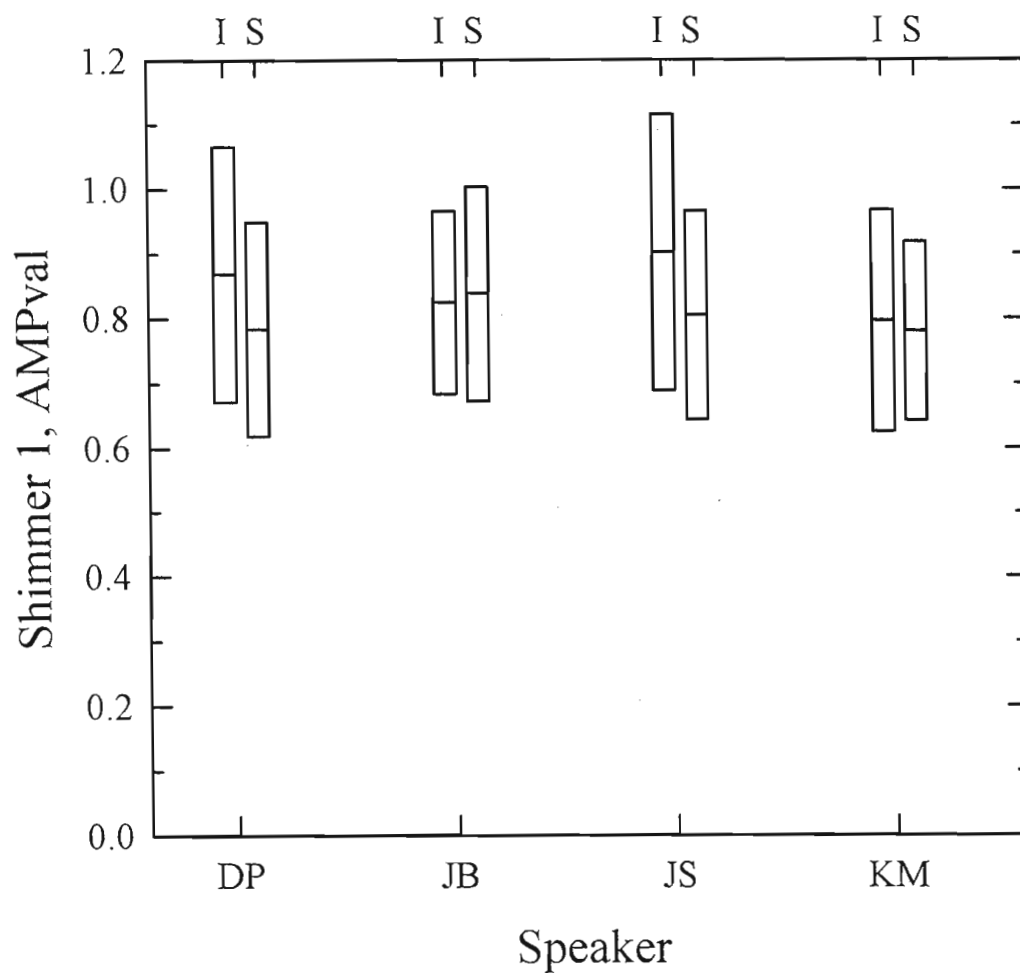


Figure 15: Sample distributions (mean and \pm one standard deviation) for 4 speakers in intoxicated (I) and sober (S) conditions for perturbation measure Shimmer 1, AMPval.

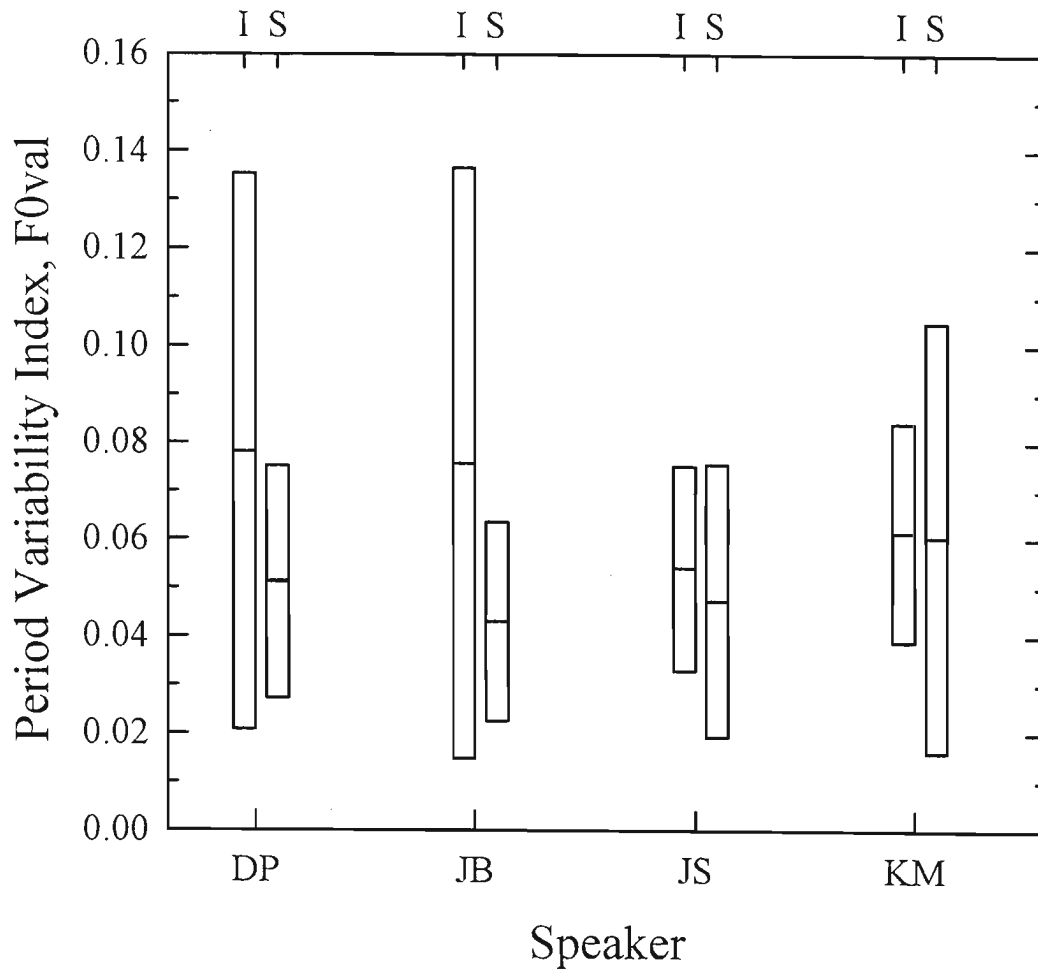


Figure 16: Sample distributions (mean and \pm one standard deviation) for 4 speakers in intoxicated (I) and sober (S) conditions for perturbation measure Period Variability Index, F0val.

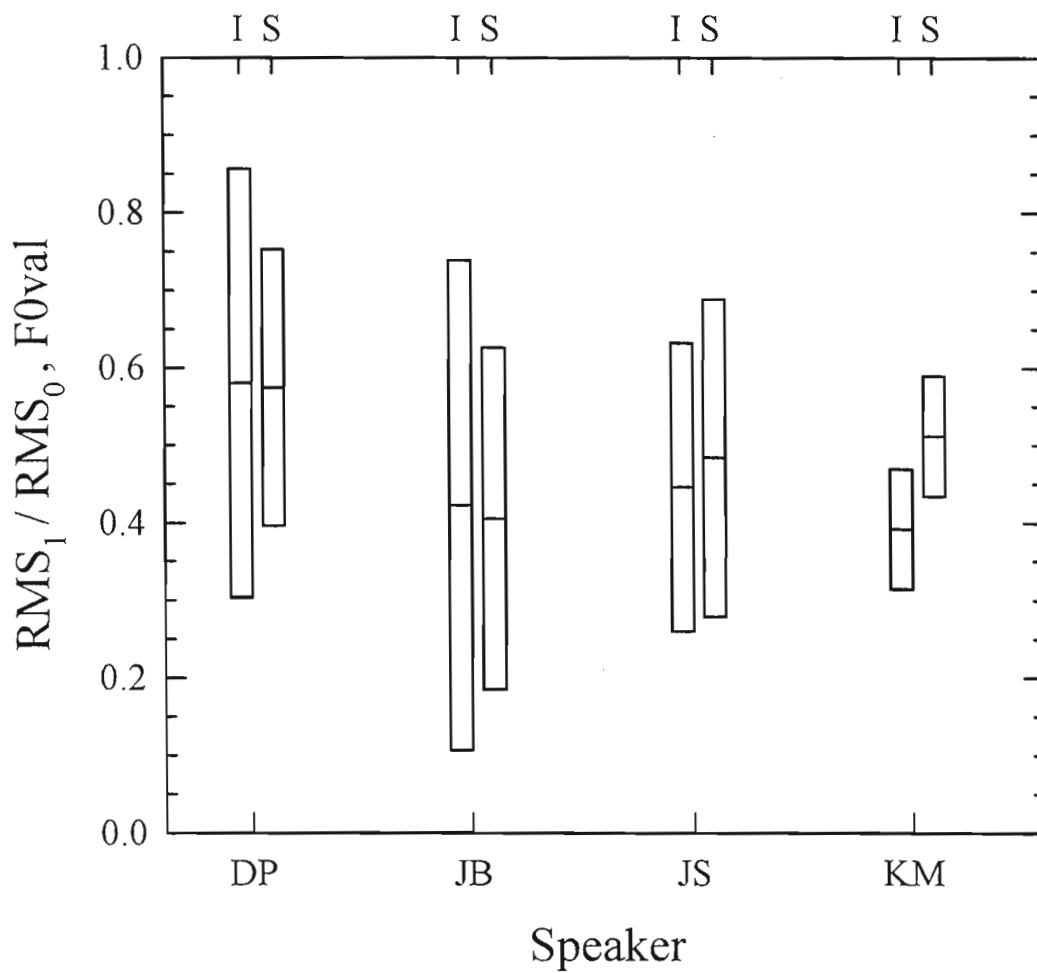


Figure 17: Sample distributions (mean and \pm one standard deviation) for 4 speakers in intoxicated (I) and sober (S) conditions for perturbation measure $RMS_1/RMS_0, F0val$.

analysis error waveform. Sample-by-sample covariance LP modelling is performed on the speech waveform using a segment of speech that is shorter than the expected length of glottal closure. When the error is closest to zero, the speech segment most closely resembles an all-pole model. This segment is assumed to include only the vocal tract resonances; thus, the glottis is closed during this segment.

Once a segment of speech during which the glottis is closed is identified, a vocal tract model is determined using 12-pole covariance LP analysis. Ten of the poles represent the ten poles of the vocal tract model. The other two are the poles of the two-pole, two-zero model of radiation at the lips. The speech is inverse-filtered with the 12-pole model and with a two-zero filter representing the two zeros in the radiation model. The resulting waveform is assumed to be a representation of the glottal excitation waveform.

The z-Transform representation is

$$G(z) = \frac{S(z) \cdot (1 - \sum_{k=1}^{10} a_k z^{-k})(1 - \sum_{l=1}^2 \alpha_l z^{-l})}{1 - \sum_{j=1}^2 \beta_j z^{-j}}$$

where

a_k = LP coefficients of the vocal tract model

$\alpha_l \beta_l$ = coefficients of the polynomials representing the two-zero, two-pole model of radiation at the lips

$S(z)$ = z-Transform of the windowed speech signal

$G(z)$ = z-Transform of the glottal excitation waveform.

Some examples of extracted glottal waveforms are shown in Figure 18. Several qualitative observations can be made concerning the differences between sober and intoxicated glottal excitation waveforms, as indicated in the following section.

 Insert Figure 18 about here

Summary of Results. Results from the glottal excitation waveshape analysis for two talkers were as follows.

- The vocal tract is less stationary in intoxicated speech than in sober speech. The LP vocal tract model can only be used to inverse-filter about four pitch periods of intoxicated speech. After that, the model is no longer accurate enough to generate a "clean" glottal waveform. With sober speech, a given LP model accurately represents the vocal tract for several pitch periods (often as many as eight or nine).

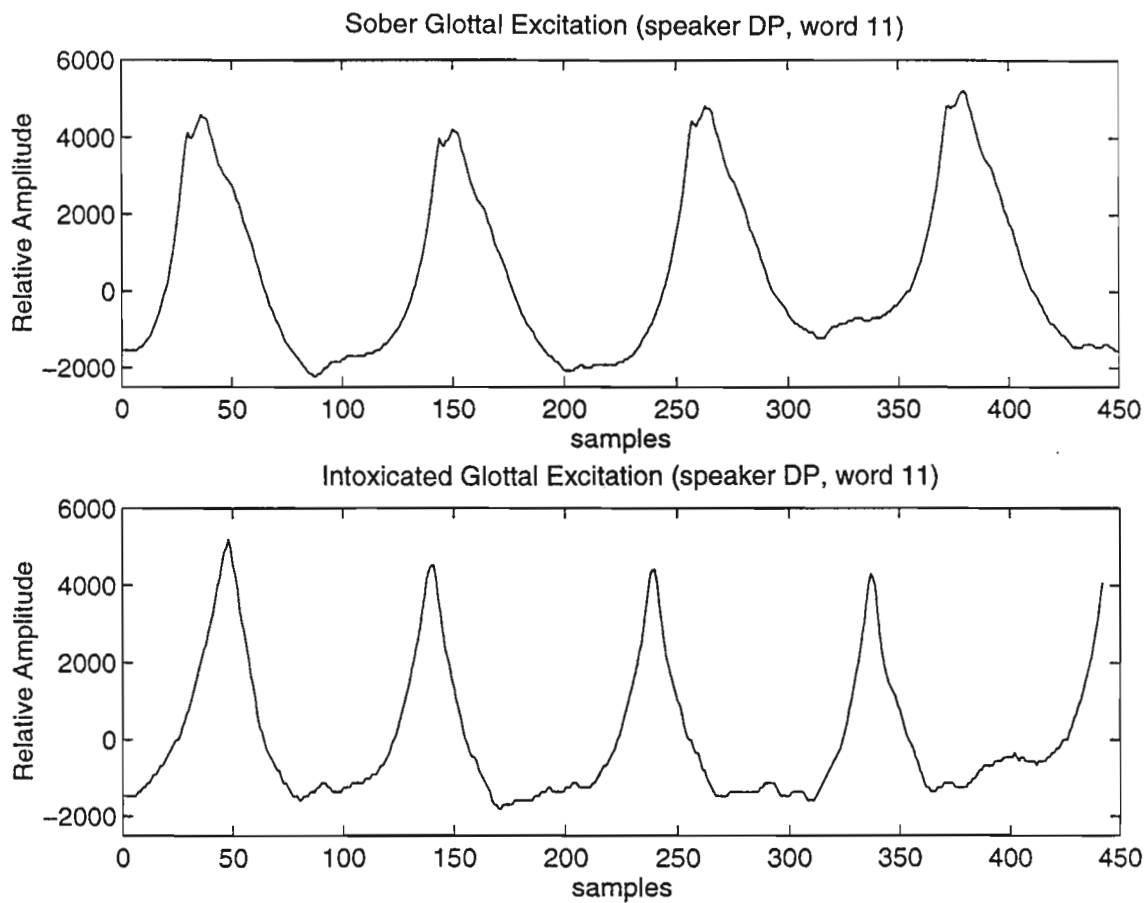


Figure 18: Examples of extracted glottal waveforms.

- The glottal waveshape is less consistent through an utterance in intoxicated as opposed to sober speech. More variance in amplitude, slopes, durations, and overall waveshape are seen in the intoxicated glottal waveforms than in the sober glottal waveforms.
- The glottal pulse shape appears to be more symmetric and more triangular in shape for intoxicated speech than for sober speech.
- Glottal closure appears to have more ripple in intoxicated speech than in sober speech.

Future Work

The work remaining on this project include the following.

- Direct measures of the acoustic speech waveform:

Extract and analyze these parameters for the remaining five speakers in the alcohol database.

- Perturbation measures of the acoustic speech waveform:

Add a shimmer parameter based on the peak amplitude within a pitch period.

Extract and analyze the perturbation measures for the remaining five speakers in the database.

- Glottal excitation waveshape:

Continue to extract glottal waveforms from all nine speakers (do the four speakers DP, JB, JS, and KM first).

Parameterize and mark all glottal waveforms for analysis.

Calculate and compare the cepstra of sober versus intoxicated speech.

- Statistical Analysis:

Based on means and standard deviations, perform statistical analysis to determine which parameters are significantly different in a statistical sense.

Develop a parameter set that can be used to distinguish between sober and intoxicated speech.

References

- Baken, R.J. (1987). *Clinical Measurement of Speech and Voice*. Boston, MA: College-Hill Press.
- Cummings, K.E. (1992). *Analysis, Synthesis, and Recognition of Stressed Speech*. Ph.D. dissertation, Georgia Institute of Technology.
- Hollien, H., Michel, J., & Doherty, E.T. (1973). A method for analyzing vocal jitter in sustained phonation. *Journal of Phonetics*, 1, 85-91.
- Jacob, L. A. (1968). A normative study of laryngeal jitter. Master's thesis, University of Kansas.
- Koike, Y. (1973). Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Studia Phonologica*, 7, 17-23.
- Ludlow, C., Coulter, D., & Gentges, F. (1983). The differential sensitivity of measures of fundamental frequency perturbation to laryngeal neoplasms and neuropathologies. In D.M. Bless & J.H. Abbs (Eds.), *Vocal Fold Physiology: Contemporary Research and Clinical Issues* (Chap. 33, pp. 381-392). San Diego, CA: College-Hill Press.
- Pinto, N. B., & Titze, I. R. (1990). Unification of perturbation measures in speech signals. *Journal of the Acoustical Society of America*, 87, 1278-1289.
- Pisoni, D.B., & Martin, C.S. (1989). Effects of alcohol on the acoustic-phonetic properties of speech: Perceptual and acoustic analyses. *Alcoholism: Clinical and Experimental Research*, 13, 577-587.
- Pisoni, D.B., Yuchtman, M., & Hathaway, S.N. (1986). Effects of alcohol on the acoustic-phonetic properties of speech. In *Alcohol, Accidents, and Injuries* (pp. 131-150). Warrendale, PA: Society of Automotive Engineers.
- Starmer, G. A. (1989). Effects of low to moderate doses of ethanol on human driving-related performance. In K.E. Crow & R.D. Batt (Eds.), *Human Metabolism of Alcohol, Volume I: Pharmacokinetics, Medicolegal Aspects, and General Interest* (pp. 101-130). Boca Raton, FL: CRC Press.
- Wong, D., Markel, J., & Gray, A., Jr. (1979). Least squares glottal inverse filtering from the acoustic speech waveform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-27, pp. 350-355.