Effects of Alcohol on the Acoustic-Phonetic Properties of Speech: Perceptual and Acoustic Analyses

David B. Pisoni and **Christopher S. Martin**
Indiana University, Bloomington, Indiana 47405

Abstract

This report summarizes the results of a series of studies that examined the effects of alcohol on the acoustic-phonetic properties of speech. Audio recordings were made of male talkers producing lists of sentences under a sober condition and an intoxicated condition. These speech samples were then subjected to perceptual and acoustic analyses. In one perceptual experiment, listeners heard matched pairs of sentences from four talkers and were required to identify the sentence that was produced while the talker was intoxicated. In a second perceptual experiment, Indiana State Troopers and college undergraduates were required to judge whether individual sentences presented in isolation were produced in a sober or an intoxicated condition. The results of the perceptual experiments indicated that groups of listeners can significantly discriminate between speech samples produced under sober and intoxicated conditions. For acoustic analyses, digital signal processing techniques were used to measure acoustic-phonetic changes that took place in speech production when the talker was intoxicated. The results of the acoustical analyses revealed consistent and well-defined changes in speech articulation between sober and intoxicated conditions. Because speech production requires fine motor control and timing of the articulators, it may be possible to use acoustic-phonetic measures as sensitive indices of sensory-motor impairment due to alcohol consumption.

Alcohol is generally considered to be a central nervous system depressant. Medium and high blood concentrations of alcohol have been found to impair intellectual functioning, reaction time, coordination, reflexes, and nerve transmission.[1] Alcohol consumption is also thought to produce changes in speech production that are often described as "slurred speech." Changes in speech production after alcohol consumption are often used by law enforcement personnel, bartenders, and others as general indices of motor impairment due to alcohol intoxication. Changes in speech production have also been used as a sign of impairment due to other drugs. Shagass[2] related slurring of speech to the initial threshold of consciousness impairment produced by sodium amythal. Very little research, however, has explored the nature of acoustic-phonetic changes in the speech waveform due to alcohol intoxication. Because speech production requires fine motor control, timing, and coordination of the articulators, it may be possible to use acoustic-phonetic measures as sensitive indices of impairment due to alcohol intoxication.

Some research relevant to this problem has been conducted. Several studies have examined the general effects of alcohol on speech production. Moskowitz and Roth[3] examined the effects of alcohol on response latency in a picture-naming task. Thirty pictures of words chosen from a word frequency list were named by 12 subjects while sober and after consuming a beverage designed to achieve a blood alcohol concentration (BAC) in the range

of 0.06–0.08%. The researchers found that alcohol increased response latency, especially for the less frequently used words. Andrews, Cox, and Smith[4] administered a moderate dose of alcohol to subjects and recorded samples of their speech. Raters, unaware that some of the recordings were produced in an intoxicated condition, listened to samples of the subjects' sober and intoxicated speech. Speech samples produced by subjects while they were sober were rated as coming from more efficient, reasonable, self-confident, scholarly, artistic, theatrical, and less untrained people than those produced by subjects while they were intoxicated.

Sobell and Sobell[5] had 16 male alcoholics read a passage while sober, after ingesting 5 oz of 86 proof alcohol, and again after ingesting 10 oz of 86 proof alcohol. At high doses of alcohol, the subjects took longer to read the passage and had more word interjections, phrase interjections, sound interjections, word omissions, word revisions, and broken suffixes in their speech. In a later follow-up study, Sobell and Sobell[6] examined the effects of alcohol consumption on the speech of nonalcoholics. Sixteen adult male talkers read a passage while sober, after receiving a dose of alcohol designed to raise their BAC to 0.05%, and again after receiving a dose of alcohol designed to raise the BAC to 0.10%. They found that the amplitude of speech decreased as blood alcohol level increased. In addition, reading rate was slower after subjects had received the high dose of alcohol than when they were sober or had received the moderate dose of alcohol. No significant effect on fundamental frequency (vocal pitch) was obtained.

Several studies have examined the effects of alcohol-intoxication on articulatory control in greater detail. Trojan and Kryspin-Exner[7] had three subjects name pictures and speak spontaneously while sober and at two levels of intoxication. They found that subjects were more likely to make sentence level, word level, and sound level errors when intoxicated. The phonemes /l/, /r/, /s/, /ʃ/, and /ts/ were the most affected by the consumption of alcohol. The effects of alcohol intoxication on pitch varied from speaker to speaker and no general pattern emerged from the analyses. In an acoustic-phonetic study conducted by Lester and Skousen,[8] a small group of subjects read prepared word lists and were engaged in conversation at various points during a gradual loss of sobriety. These investigators found that as subjects became more intoxicated, they showed an increased tendency to lengthen consonantal segments in unstressed syllables, devoice word-final obstruents, and retract the place of articulation for /s/. Deaffrication of /tʃ/ and /dʒ/ also occurred in their speech.

In summary, speech produced under intoxication has been found to be slower, lower in overall amplitude, more negatively judged in subjective perceptual tests, and more prone to errors at the sentence, word, and phonological levels than speech produced in a sober condition. The nature of the sound errors reported in several of the studies cited above suggests that alcohol reduces the control and coordination of speech articulation, phonation, and respiration, particularly the fine motor control required for the articulation of consonants such as stops, fricatives, and affricates.

Except for these fairly general observations, very little quantitative data is currently available in the published literature on the effects of alcohol on speech production, particularly in terms of the acoustic-phonetic characteristics of speech. Many of the studies in the literature suffer from methodological problems. For example, most of these studies did not objectively measure the BACs of their subjects after ingestion of alcohol. Measurements of the speech samples, when they were made, used fairly gross analog techniques. As far as we have been able to determine, no studies have applied modern digital signal processing techniques, which allow more precise acoustic measurements. Finally, except for the study by Lester and Skousen,[8] no efforts have been made to examine detailed

changes in the acoustic-phonetic properties of speech produced by talkers in an intoxicated condition.

It is obvious that a well-designed laboratory investigation is needed in this area. In the present investigation, speech samples produced by male talkers while sober and after obtaining a BAC at or above 0.10% were subjected to both perceptual and acoustic analyses. Objective quantitative procedures for assessing the BACs of the talkers were employed. The speech samples were used in perceptual experiments in order to examine the ability of listeners to reliably judge whether speech was produced under intoxication. In addition, acoustic-phonetic measures of the changes in speech production were obtained using digital signal processing techniques.

SPEECH PRODUCTION

Talkers

Eight male students enrolled at Indiana University were recruited through a newspaper advertisement and were paid to serve as talkers in a two-part experiment. All subjects were at least 21 years of age, native speakers of English, and had no history of a speech, language, or hearing disorder at the time of testing. Each subject completed an alcohol consumption questionaire, the short Michigan Alcoholism Screening Test,[9] the MacAndrew scale,[10] and the socialization subscale of the California Psychological Inventory.[11] These tests are frequently used to identify subjects who are theoretically at risk for the development of alcoholism. The MacAndrew scale has been found to predict future alcoholism in nonalcoholics.[12] The socialization subscale of the CPI was designed to measure a constellation of personality traits that have been found to predict future alcoholism.[13] Only subjects whose scores on these tests showed them to be moderate social drinkers at low risk for alcoholism were included in the experiment. Subjects were required to abstain from food and drink for at least 4 hr prior to the experiment.

Materials

Auditory stimuli were used in the experiment to elicit samples of speech from the talkers. The auditory stimuli consisted of 66 sentences spoken in citation format by a male talker. These stimuli were chosen to present varying degrees of articulatory difficulty.[14] All auditory stimuli were first prerecorded in a sound-attenuated IAC booth using an Electro-Voice Model D054 microphone and an Ampex AG-500 tape recorder. The stimuli were then low-pass filtered at 4.8 kHz and digitized at a 10-kHz sampling rate through a 12-bit A/D converter. A digital waveform editor[15] was used with a PDP 11/34 minicomputer to edit all speech samples into separate digital files for later playback and recording of test tapes. Four audio tapes were produced using a computer-controlled audio tape-making program. The digital waveforms were output through a 12-bit D/A converter, low-pass filtered at 4.8 kHz and recorded on audio tape at a speed of 7.5 inches per second. All of the sentences were then recorded with 3 sec of silence after each sentence. A different random ordering of the sentences was used for each tape.

Procedure

Talker Preparation—Each talker was seen individually and participated in two sessions. During one session, the subject was sober; during the other session, the subject consumed enough alcohol to achieve a BAC at or above 0.10% weight/volume. Sessions were counterbalanced to control for order effects. Audio recordings were made at each session. A Smith and Wesson Breathalyzer (Model 900A) was used to measure subjects' BACs. In the sober condition, the talker was given a breath analysis test prior to having his speech recorded to insure that there was no alcohol in his system. In the intoxicated condition, the

talker was weighed and given a breath analysis test. A mixture of 1 part 80 proof vodka and 3 parts orange juice was prepared using 1 g of alcohol for every kilogram of the talker's weight. This dose was designed to raise the talker's BAC to 0.10%. The talker was given one-third of the total dose every 15 min and was told to pace his drinking to cover the entire 15 min period. At the end of 45 min, the talker was asked to rinse his mouth several times to remove all traces of alcohol from his mouth. After waiting 5 min, talkers were given a breath-analysis test. If a talker's BAC was still below 0.10%, he was given another drink containing the same amounts of vodka and orange juice as each of the previous three drinks. This drink was consumed over a period of 15 min. The subject then repeated the mouth-rinsing, 5-min wait, and breath-analysis test. When the talker's BAC was at or above 0.10%, recordings of his speech were made. Table 1 shows each talker's BACs prior to the recording session.

Recording—Talkers sat in a single-walled sound-attenuated IAC booth and wore a pair of matched and calibrated TDH-39 headphones with an attached EV C090 LO-Z condenser microphone throughout the experiment. The microphone was mounted on a boom and was adjusted so that its placement was 4 inches directly in front of the subject's mouth.

Talkers remained naive to the real purpose of the experiment. They were told that the experiment involved the effects of alcohol on memory and the rate at which talkers could read and shadow material while under the influence of alcohol. All audio recordings of the talkers' speech were made with an Ampex AG-500 tape recorder. Recording levels were adjusted at the beginning of a subject's first session and, in order that amplitude could be compared across conditions, remained the same throughout both sessions of the experiment.

Auditory stimuli for shadowing were presented for all subjects via audio tape playback over headphones. Subjects were instructed to listen carefully to each sentence and then to quickly repeat it back (shadow) aloud as soon as they could. Following completion of the experiment, subjects in the intoxicated condition were given a final breath-analysis test and were then sent home in a prepaid taxi.

PERCEPTUAL EXPERIMENTS

EXPERIMENT 1

Method

Listeners: Listeners were 21 undergraduate students recruited from an introductory psychology course who participated in the experiment to fulfill part of a course requirement. Each listener participated in one 45-min session. No listener reported any history of a speech or hearing disorder at the time of testing.

Materials: Thirty-four pairs of shadowed sentences, recorded in both the sober and intoxicated conditions by each of four talkers (Talkers 1 through 4 in Table 1), were selected for use as test stimuli in the perceptual experiment. These 34 sentences, each containing one or two "key words," were used previously in nerve-block studies carried out by Borden and her associates.[12,14,15] The key words in these sentences contained a wide range of phonemes and placed special emphasis on fricative clusters.

The samples of speech produced by each talker were low-pass filtered at 9.6 kHz and digitized at a 20-kHz sampling rate through a 12-bit A/D converter. Using a digital waveform editor, the 34 test sentences were extracted from the audio recordings for each talker in both conditions. Two sets of matched pairs of sentences (A and B) were created by digitally splicing each sentence spoken by a talker in the sober condition with the same sentence spoken by the same talker in the intoxicated condition. In the "A" set, the order of

the sentences in each pair (either sober-intoxicated or intoxicated-sober) was randomly assigned for each sentence spoken by each talker. In the "B" set, the ordering of sentences within each pair was the opposite of the ordering used in the "A" set. In both sets of sentence pairs, each sentence was separated from the other by one second of silence.

Two test tapes were made using sentences from the four talkers. On one tape, pairs of sentences from the "A" set were recorded in random order. On the other tape, pairs of sentences from the "B" set were recorded in random order. On both test tapes 4 sec of silence were inserted between pairs of sentences. Sentences with a frequency cutoff of 4.9 kHz were used for Talkers 1 and 2; sentences with a frequency cutoff of 9.8 kHz were used for Talkers 3 and 4.

Procedure: Listeners were asked to listen carefully to each pair of sentences and to decide which sentence in the pair (i.e., "A" or "B") was produced by the talker while intoxicated. Listeners indicated their choice by circling "A" if they thought that the first sentence in the pair was produced while the talker was intoxicated or "B" if they thought that the second sentence was produced while the talker was intoxicated. Listeners were informed that one and only one of the sentences in each test pair was produced by the talker while intoxicated. Half of the listeners heard one of the test tapes prepared for the experiment, and the other half of the listeners heard the second test tape.

The listeners also gave a confidence rating for each of their responses, indicating how certain they were on a scale from 1 ("just guessing") to 5 ("very sure") that they had correctly chosen the sentence produced while the talker was intoxicated.

Results—The mean proportion of correct responses across all listeners and talkers was 73.8%. Table 2 shows the proportion of correct responses made by listeners for each of the four talkers. Tests of proportions[16] showed that listeners were able to judge which sentence in each pair was produced while the talker was intoxicated at above chance performance for each of the four talkers ($p < 0.001$ for each talker). Naive listeners were able to reliably identify, at above chance levels, the sentence in each sentence pair that was produced while the talkers were intoxicated. The percentage of correct responses ranged from 70 to 80%, suggesting individual differences among the talkers in the degree of impairment in speech production at similar BACs. The discrimination performance of the listeners, while statistically better than chance, contained a substantial number of errors.

Confidence ratings given by listeners for each of their responses were positively correlated with response accuracy. The more confidence a listener placed in a given response, the more likely it was that the response was correct.

EXPERIMENT 2

The results of the first perceptual experiment demonstrated that naive listeners can reliably identify speech produced in an intoxicated state when given matched pairs of sentences produced by the same talker in a sober and an intoxicated condition. This paired-comparison procedure, however, is not the best experimental analog for clinical and law enforcement settings in which an immediate absolute judgment concerning intoxication is made on the basis of an individual's speech. An absolute identification task, in which listeners make a judgement based on a sample of speech produced in only one condition, serves as a closer experimental analog to examine listeners' ability to judge impairment via speech in a field setting.

For this reason, a second perceptual experiment was conducted using an absolute identification task. Subjects were required to make a judgment for each sentence presented

in isolation, rather than a comparative judgement on a matched pair of sentences from the same talker. Indiana State Trooper and naive student listeners were used to assess differences in discriminating speech produced in the two conditions as a function of prior experience in identifying intoxicated individuals.

Method

Listeners: Two groups of listeners were used. One group consisted of 30 introductory psychology students who received credit to fulfill a course requirement. The second group consisted of 14 Indiana State Troopers from the Bloomington post who volunteered their time to participate in this study. Each listener participated in one hour-long session. All listeners were native English speakers with no reported history of a speech or hearing disorder at the time of testing.

Materials: Twenty-four shadowed sentences produced in sober and intoxicated conditions by each of the eight original talkers were used as test stimuli in this perceptual experiment. These 24 sentences were a subset of the original 34 sentences used in the first perceptual experiment. Two master digital files of these recorded sentences were compiled. Each file contained eight talkers speaking the same 24 sentences. Each talker contributed 12 sentences produced in the intoxicated condition, and 12 different sentences produced in the sober condition. The two master files differed in that each sentence-speaker combination appeared only in the intoxicated condition for one file and the sober condition for the other. Different random orders of each master file were then transferred from the computer onto audio tapes using a 12-bit digital-to-analog converter. A 5-sec silent interval was inserted between successive sentences. Half of the listeners in each group heard an audio tape generated from the first master file, and half heard an audiotape generated from the second master file.

Procedure: Each listener heard eight talkers saying 24 sentences each, for a total of 192 sentences. Listeners wore headphones and were presented with the sentences on audio tape. They recorded their decision after each sentence by circling the letter S for "sober" or I for "intoxicated" on a prepared response sheet. Listeners then rated their degree of confidence in their choice on a scale from 1 ("just guessing"), to 5 ("very sure").

Results

Accuracy: Mean accuracy across all of the sentences was 61.5% for the college students, and 64.7% for the State Troopers. Tests of proportions demonstrated that both groups performed significantly better than chance ($p < 0.001$). Mean accuracy for the sentences produced in the two conditions was 60.5% for sober, and 64.5% for intoxicated.

Mean accuracy for the different talkers ranged from 55% to 71.9%. While tests of proportions demonstrated that accuracy for all of the talkers was significantly better than chance ($p < 0.01$ for each talker), the discrimination performance of the listeners was far from perfect, and involved a substantial number of errors. In addition, there were substantial individual differences in the speech samples produced by the eight talkers that affected the identification performance of the listeners. A series of *t* tests showed that State Troopers performed significantly better than college students for six of the eight talkers ($p < 0.05$). The difference between Troopers and students, while statistically significant for six of the talkers, was not very large. The mean percentage of correct responses by talker for the two groups of listeners is shown in Fig. 1.

Signal Detection Analysis: The percentage of correct identifications for any given talker can be misleading, because performance is affected by listeners' response biases. In order to assess listeners' discrimination performance with response bias controlled, a signal detection

analysis of the data was performed. Mean beta and d' values for the two groups were calculated from the proportion of hits (correctly identifying a sentence produced in the intoxicated condition) and false alarms (identifying a sentence actually produced in the sober condition as having been produced in the intoxicated condition). Beta is a measure of response bias, and d' is a measure of discrimination performance independent of response biases. Beta was slightly higher for State Troopers compared to students, but this difference was not significant. However, State Troopers displayed a significantly higher d' (0.786) compared to students (0.603). A *t* test demonstrated that this difference was significant ($p < 0.01$). Thus, State Troopers were better than college students in discriminating sentences produced in a sober and an intoxicated condition. This difference is probably due to the extensive experience State Troopers have in identifying intoxicated individuals.

The proportion of hits and false alarms for both groups of listeners for individual talkers is plotted in an ROC graph in Fig. 2. Performance for the individual talkers was very similar for both groups, although the State Troopers displayed a stricter criterion in judging whether a sentence was produced in the intoxicated condition for seven of the eight talkers. As shown in Fig. 2, there was considerable variability in the hit and false alarm rates for the different talkers. The speech of some talkers was consistently labeled sober or consistently labeled intoxicated in both conditions.

Confidence Ratings: Percentage accuracy across the five confidence rating categories is shown in Table 3. The more confidence a listener placed in a given response, the more likely it was that the response was correct. This finding suggests that listeners are able to reliably monitor their ability to discriminate between the speech samples produced in the two conditions.

MEASURES OF SPEECH PRODUCTION

PHONETIC TRANSCRIPTIONS OF KEY WORDS

Method—Two experienced phoneticians independently made narrow phonetic transcriptions of the key words in the test sentences spoken by four talkers (Talkers 1–4 in Table 1) in both conditions. The phoneticians listened to one of the tapes used in the second listening experiment and were unaware of which sentence in each pair was produced by a talker while intoxicated. Both transcribers listened to the tape twice. Following the first listening session, disagreements in transcription responses were marked and received special attention during a second listening session. No further attempt was made to resolve disagreements.

Results—Interjudge reliability for the transcriptions of the key words was calculated by finding the percentage of key words (produced by all four talkers in both conditions) for which the transcriptions of both phoneticians agreed after the second listening session. Transcriptions were identical for 296 of the 304 words. Thus, point-by-point agreement was approximately 97%. Errors or changes in pronunciation noted by only one phonetician were not included in the following results.

Table 4 gives a summary of the major types of errors and changes in pronunciation found in the key words of the test sentences. Overall, the four talkers made 41 changes in pronunciation after alcohol consumption, compared to only two changes in pronunciation while sober. A *t* test for correlated samples showed that significantly more errors were made in the intoxicated condition compared to the sober condition ($p < 0.01$). Vowel lengthening was the most common type of phonetic change found in the key words produced by talkers after consuming alcohol. Vowel lengthening was observed in 14 of the key word tokens produced by talkers while intoxicated, and in none of the tokens produced by the same

talkers while sober. Consonant deletion or the partial articulation of a consonant was the next most frequent type of change observed. This occurred in 13 of the tokens produced in the intoxicated condition and in one of the tokens produced in the sober condition. The liquid consonants /l/ and /r/ were the most frequently deleted or partially articulated phonemes, followed by the nasal consonants / n/ and /ŋ/ and the stop consonant /d/.

Deaffrication was another relatively common speech error found in key words produced in the intoxicated condition. Six instances of deaffrication were observed in these tokens. In three of these cases, deaffrication of the voiced affricate /dʒ/ took place; in the other three cases, deaffrication of the unvoiced affricate /tʃ/ occurred. Deaffrication of /ts/, sometimes considered a voiceless affricate, occurred in one token produced in the sober condition.

As Table 4 shows, several of the talkers were more likely to produce a distinctive change in pronunciation than others. Talker 3 produced the greatest number of errors and pronunciation changes (17), eight of which involved vowel lengthening. Talker 2, however, produced only six errors.

The results of the phonetic transcriptions indicate that, although talkers do not produce a large number of phonetic errors while intoxicated, they are more prone to produce such errors than when they are sober. The errors and changes in pronunciation found in the key words produced in the intoxicated condition suggest that the effects of alcohol reduce the talker's speaking rate by prolonging vowels and consonants. Intoxicated talkers may sometimes produce consonants with an altered place of articulation. Finally, it appears that intoxicated talkers have some difficulty with timing and coordination in the production of affricates. In considering these findings, it should be kept in mind that phonetic transcriptions are subjective perceptual judgements by trained phoneticians about the speech of a talker. These judgements cannot be objectively quantified without making physical measures of the speech waveform through acoustic analyses. These types of analyses were also carried out on the speech samples and are described in the next sections.

ACOUSTIC ANALYSES

The perceptual studies reported above indicate that groups of trained and naive listeners can reliably discriminate between sentences produced under sober and intoxicated conditions. Moreover, although the speech of intoxicated talkers had a higher frequency of phonetic errors, it is clear that, in most cases, far more subtle acoustic changes were available to the listeners in their decision making. In this section, we report the results of acoustic analyses that were carried out to measure the physical attributes of several selected speech sounds produced in sober and intoxicated conditions. Two types of analyses were carried out. First, a sentence level analysis was undertaken in which global measures of duration, overall energy, and voicing were obtained. Second, a detailed segmental analysis was also undertaken in which the acoustic properties of several classes of speech sounds were measured in both the frequency and time domains. The goal of these acoustic analyses was to quantify the impressions obtained from phonetic transcriptions and to verify the perceptual findings obtained from the listening experiments.

Methods and Procedures

Speech Materials: The materials for the acoustic analyses consisted of the same sentences used in the first perceptual study, namely, the 34 sentences spoken by each of four talkers (Talkers 1–4 in Table 1) under both sober and intoxicated conditions. Analysis of the fine acoustic-phonetic structure of the speech was performed on segments derived from the 38 keywords transcribed by the trained phoneticians.

Digital Signal Processing: All sentences were low-pass filtered (9.6-kHz cut-off) and digitized at a rate of 20,000 samples/sec. Each digitized sentence was stored in a computer file for further processing. Preliminary analysis of the sentences involved application of several digital signal processing algorithms selected from the ILS (Interactive Laboratory System) software package. Linear prediction analysis (autocorrelation method) was performed using a window length of 25.6 msec, with a shift ("frame") of 12.8 msec between consecutive windows. Prior to computation, each analysis window was filtered (Hamming window) and preemphasized. The LPC analysis yielded 21 coefficients from which formant frequencies, bandwidths, and amplitudes, as well as overall power level, were calculated. In addition, a pitch extraction algorithm was employed to determine if a given segment was voiced or unvoiced and, for voiced segments, to estimate the fundamental frequency (vocal pitch).

Sentence Analysis: The purpose of the sentence-level analysis was to evaluate the effects of alcohol intoxication on several acoustic parameters calculated for the duration of whole sentences, thus reflecting the global or long-term properties of the speech. The acoustic parameters that we examined fell into one of three categories:

1. Temporal Measures

   a   Overall sentence duration

   b   Total duration of unvoiced segments of the sentence

   c   Total duration of voiced segments

   d   Voiced-to-unvoiced ratio

2. Voicing Decision Parameters

   e   First reflection coefficient

   f   Value of zero-crossing

   g   Amplitude of cepstral peak

   h   Amplitude of absolute and normalized residuals

   i   Voicing decision statistic (a value representing a weighted combination of parameters e and f)

3. Pitch Measures

   j   First four moments of FO

Segmental Analysis: The segmental analysis was carried out in order to evaluate the effects of alcohol on the acoustic properties of specific classes of speech sounds. Each of the 38 keywords was divided into a sequence of labeled segments (i.e., phonemes). The labels included the following segment types:

1. Voiced strident fricatives (z, ʒ)

2. Unvoiced strident fricatives (s, ʃ)

3. Voiced weak fricatives (v, ð)

4. Unvoiced weak fricatives (f, ɵ)

5. Voiced affricates (dʒ)

6. Unvoiced affricates (tʃ)

7. Stops

8. Closures preceding voiced stops

9. Closures preceding voiceless stops

10. Nasals

11. Vocalic (vowels, glides, and liquids)

The segmentation process was performed by visual inspection of the speech waveforms using an interactive digital processing program with cursor controls on a CRT graphics workstation interfaced to a DEC VAX 11/750 computer. Spectral information as well as power and voicing parameters were displayed simultaneously on the CRT to facilitate exact definitions of segment boundaries. Following segmentation and labeling of a given segment, a program computed, stored and labeled each of the 38 keywords. The program also calculated for each segment all of the acoustic parameters listed above. In addition, several spectral measures were calculated. These included the slope of the mean power spectrum of the segment, the half-power frequency (the point above and below which half of the energy is concentrated) and, finally, normalized distance measures between the onset, middle, and offset of the segment. Mean values and variances of the acoustic parameters were calculated for all segment categories.

Results

Sentence-Level Analysis: Of all the acoustic measures, an overall change in duration was the most consistent difference observed between speech samples produced in the two conditions. For all four talkers, the mean duration of their sentences was consistently longer when produced in the intoxicated condition. The average magnitude of sentence lengthening for individual talkers ranged from 75 to 158 msec.

Figure 3 shows the proportion of sentences produced in the intoxicated condition that were lengthened or shortened relative to the same sentences produced in the sober condition. For individual talkers, between 70 to 97% of the sentences had longer durations in the intoxicated condition. Sign tests performed on the overall duration measures demonstrated that alcohol-intoxication reliably increased the duration of the sentences for each of the four talkers ($p < 0.05$).

No consistent changes were observed in the acoustic measures that reflect laryngeal control of voicing or pitch. Contrary to what may be an intuitive hypothesis regarding the effects of alcohol on vocal level, only slight differences were observed in the amplitude of both voiced and voiceless segments in the speech produced in the two conditions. As shown in Fig. 4, the pitch level of all four talkers was, however, much more variable in the intoxicated condition, suggesting less precise control of the rate of vocal cord vibration. A *t* test for correlated samples revealed that pitch variability was greater in the intoxicated condition compared to the sober condition. The mean values of pitch, however, changed only slightly for three of the four talkers. The pitch of Talker 3 was considerably lower in the intoxicated condition compared to the sober condition. This decrease was observed for this talker in all voiced segments and throughout the sentences, suggesting a dramatic change in his control of fundamental frequency in speech.

Segmental Analysis: An analysis of selected phonetic segments was carried out in order to identify the extent to which the articulation and timing of specific speech sounds was affected by alcohol intoxication. Changes in articulation would be reflected in the spectral properties of these sounds. In the present study, analyses of fricatives, stops, nasals, and vowels did not reveal any consistent spectral differences which could be attributed to the

effects of alcohol. The locus of the effect of alcohol was identified by an analysis of various segments in the time domain rather than in the spectral domain. Several classes of speech sounds were affected, however, in terms of the timing of particular articulatory events. These changes are summarized below.

The production of a stop consonant sound typically involves a sequence of several articulatory movements consisting of a very brief and complete obstruction of the vocal tract at the lips or tongue, followed by an abrupt release of energy. The release of air pressure from behind the closure results in a short burst of noise followed by a period of fricative turbulence. For voiced stops, voicing (glottal vibration) may cease for a short time, or continue throughout the closure and release phases. Figure 5 shows the waveform of the /d/ portion of the word "dishes" for the four talkers in both sober and intoxicated conditions. This figure illustrates the longer closure durations typically obtained in the analyses of stop consonants after alcohol consumption. A *t* test for correlated samples revealed that talkers had considerably longer closure durations for stop consonants when intoxicated than when sober ($p < 0.01$).

The articulation of affricates was affected in a similar manner. Normally, an affricate sound is produced by a complete closure of the vocal tract followed by an abrupt release and prolonged frication. Figure 6 shows the waveform of /dʒa/ from the word "pajamas." Three of the four talkers showed incomplete closure of the vocal tract which is marked by a "leak" of noise prior to the onset of the fricative portion of the sound. The difficulty of achieving or maintaining appropriate closure in the articulation of stops and affricates is further demonstrated in Fig. 7: This waveform depicts the sounds /dʒ/ and /k/ in the words "garbage cans." The articulation of these sounds in close proximity to each other requires very precise timing and control of the degree of opening of the vocal tract. When intoxicated, talkers were not able to achieve a complete closure either before the affricate or before the stop. These observations were generally more pronounced and consistent for the articulation of voiced affricates compared to unvoiced affricates.

The production of /ts/ clusters, although often not regarded as an affricate in English, shows considerable similarity to voiceless affricates in terms of the articulatory events associated with its production. It is not surprising, then, that under the intoxicated condition the talkers also encountered the same difficulties associated with an abrupt onset and offset of closure.

The production of postvocalic /z/ is another example in which frequent departures from normal articulation were observed. Typically, in a vowel-/z/ sequence, as the amplitude of the vowel decreases, constriction of the vocal tract begins, accompanied by glottal pulsation. Finally, as maximum constriction is achieved, strong frication noise is produced accompanied by intermittent voicing. When intoxicated, talkers displayed two types of deviations from this general pattern. In several cases, the onset of constriction began well within the vowel segment; in other cases, an unvoiced interval followed the vowel, resulting in unvoiced /z/.

In summary, the global sentence-level and more detailed segmental analyses of the speech samples indicated that the acoustic properties which were most consistently affected by alcohol included the overall timing of articulation, most noticeably in the fine motor control needed for the coordination of rapid onsets and offsets of stop and affricate closures. Moreover, these effects became even more pronounced when the movement of the articulators had to be coordinated with activation or deactivation of the voicing mechanism controlled by the larynx. Thus, in those sounds requiring both precise timing and positioning of the articulators, large and reliable effects were observed. When the subjects had sufficient

time, as in the production of fricatives and other continuant sounds, they typically did not have difficulties achieving the correct place of articulation.

GENERAL DISCUSSION

The overall goal of this investigation was to determine how the acoustic-phonetic properties of speech are affected when a talker is under the influence of alcohol. The results of the perceptual experiments indicate that groups of listeners can reliably discriminate between tokens of speech produced in sober and intoxicated conditions, although listeners made a substantial number of errors. Moreover, there appears to be reliable differences between groups of listeners as a function of their experience in detecting these properties in speech. State Troopers were slightly better than college students in discriminating speech produced under these conditions, as measured by the unbiased discriminability index d'. The phonetic transcriptions of key words by trained phoneticians suggested that, while the speech of intoxicated talkers contained a higher frequency of misarticulations, the absolute frequency of salient errors was quite low and could not by itself account for the discrimination accuracy of the listeners. We conclude that speech produced under conditions of alcohol intoxication deviates from normal speech in some additional, less pronounced manner, which contributes to reliable perceptual discrimination. The present results demonstrate that regular and systematic changes in sensory-motor control due to alcohol intoxication are encoded in the speech waveform and that these changes can be reliably identified by groups of listeners. (It is possible that the degree of changes in speech produced after alcohol consumption may vary as a function of talkers' motivation to disguise any impairment. For example, the present results may not generalize to speech produced in situations in which a person is trying to appear sober before a law enforcement officer. The degree to which talkers can mask any changes in speech after alcohol consumption is currently an unanswered question.)

The acoustic analyses of the matched speech samples indicate that several acoustic-phonetic properties were consistently affected by alcohol consumption. The major global effect was restricted to the overall duration of sentences; talkers in the intoxicated condition decreased their speaking rate significantly. Beyond the duration effects, the results demonstrate that intoxicated talkers display difficulty in controlling the abrupt closure and opening of the vocal tract, especially when such articulatory gestures have to be coordinated closely with appropriate voicing behavior controlled at the larynx. Typically, this difficulty resulted in exceedingly long durations of closures before voiced stops or the complete absence of closures before affricates.

For the most part, place of articulation in the production of the target sounds was not affected by alcohol intoxication. It is interesting to note here that the effects of alcohol on speech production observed in the present investigation differ quite substantially from what is known about the effects of local nerve-block anesthesia on speech articulation.[14,17,18] The latter consists primarily of an impaired ability to maintain a precise place of articulation, most notably for sounds that require a narrowing of the vocal tract (e.g., /s/, /z/). Nerve-block anesthesia suppresses sensory input from the articulators, and thus interrupts the kinesthetic feedback which is necessary for a precise control over their placement and movement. Alcohol, on the other hand, is largely a central nervous system depressant, and the results of this study can be accounted for quite simply by this locus of effect: motor activity is slowed down and, with increasing BAC, central control and coordination of motor behavior becomes more severely impaired. These processes are reflected in the articulatory control of speech, and, in turn, in the acoustic correlates of articulatory gestures that are used to produce the sound sequences encoded in the speech waveform.

On the basis of the acoustic data, it is possible to define different levels of articulatory impairment in speech production due to alcohol intoxication. Talker 2, for example, was possibly the least affected by alcohol because he showed the smallest duration effect, both in terms of proportion of longer sentences and in terms of the magnitude of the effect. Substantial individual differences in the talkers' degree of impairment at similar BACs was also suggested by the results of the perceptual experiments. A great deal of variation was observed not only in the percentage of correct perceptual identifications given for different talkers by listeners, but also in the number of transcribed phonetic errors for key words made by the trained phoneticians.

It is well known that large individual differences exist in the degree to which persons are impaired by alcohol at similar BACs.[19] The present results suggest that measures of the acoustic-phonetic properties of speech may provide sensitive indices to quantify the degree of impairment induced by alcohol intoxication.

Several additional issues must be addressed, however, before speech production measures can be used widely as sensitive indices of sensory-motor impairment due to alcohol intoxication. First, there is the issue of the magnitude of the effects of intoxication on speech measures. The mean duration effect found in acoustic analyses of sentences is one such example. Since speech rate varies considerably among different speakers, it is quite reasonable to expect a high degree of variability and overlap. In order to define the magnitude of effects within some probabilistic framework, it will be necessary to expand and improve sampling techniques in the types of talkers used and the types of speech samples recorded. Although consistent changes were observed in the present investigation among all talkers, given the inherent variability of speech, more research will be needed to assess the probabilities of given changes as well as the contribution of general and idiosyncratic effects.

A second issue concerns the reliability of the effects of alcohol intoxication on speech production measures. Many more replications of each condition will be needed in order to establish the reliability of observed acoustic changes over time. In addition, speech materials should include several tokens of each utterance in order to define the distributional properties of the acoustic measures. Our results indicate that speech samples containing a larger number of affricates, stops and stop-consonant clusters will be more useful than the current samples in studying the effects of alcohol intoxication on speech production.

A third issue is the sensitivity of speech production changes at lower BACs and lower levels of alcohol-induced impairment. The present research investigated differences between speech produced in a sober condition and when talkers achieved BACs at or above 0.10%. Future research should examine whether reliable changes in the acoustic-phonetic properties of speech are produced when talkers are less intoxicated than in the present investigation. Speech measures should also be compared to other measures of motor or psychomotor performance such as body sway at different blood alcohol levels, in order to examine the relative sensitivity of different measures of sensory-motor impairment.[18] A related issue is whether the effects of alcohol intoxication on speech production are different on the ascending and descending limbs of the blood alcohol curve.

Finally, a fourth issue concerns the specificity of the effects of alcohol intoxication on speech production. If changes in speech production are to be used effectively as indices of sensory-motor impairment due to alcohol intoxication, these changes should not be induced by factors such as the talker's level of fatigue or stress.

Considering these limitations, the results of this investigation clearly demonstrate that alcohol intoxication affects the acoustic-phonetic properties of speech in a systematic and

consistent manner. Moreover, these changes are consistent with what is currently known about the articulation of speech sounds and the motor control mechanisms used to generate these sounds in the vocal tract (see Ref. 20).

In summary, the results of the present investigation permit us to make several general conclusions about the effects of alcohol on speech production. First, groups of listeners can reliably discriminate sentences produced in sober and intoxicated conditions. Second, very consistent acoustic-phonetic changes were observed in matched speech samples from talkers who had BACs at or above 0.10%. Third, the changes observed in speech production under alcohol intoxication occur at both the sentence level and in the fine articulation of individual speech sounds. Impaired motor control over fine articulatory movements was observed primarily in the production of voiced stops, affricates, and stop clusters. Finally, considering the well-known observations of large individual differences in the susceptibility to alcohol intoxication, differences in the acoustic-phonetic properties of speech could serve as sensitive indices of sensory-motor impairment due to the depressant effects of alcohol on central nervous system functions.

Acknowledgments

References

1. American Medical Association Committee on Medicolegal Problems. Alcohol and the Impaired Driver. Chicago: American Medical Association; 1968.

2. Shagass C. The sedation threshold: a method for estimating tension in psychiatric patients. Electroenceph Clin Neurophysiol. 1954; 6:221–233. [PubMed: 13161845]

3. Moskowitz H, Roth S. Effect of alcohol on response latency in object naming. Quart J Stud Alcohol. 1971; 32:969–975. [PubMed: 5137096]

4. Andrews ML, Cox WM, Smith RG. Effects of alcohol on the speech of non-alcoholics. Central States Speech J. 1977; 28:140–143.

5. Sobell LC, Sobell MB. Effects of alcohol on the speech of alcoholics. J Speech Hearing Res. 1972; 15:861–868. [PubMed: 4652408]

6. Sobell LC, Sobell MB, Coleman RF. Alcohol-induced dysfluency in nonalcoholics. Folia Phoniatrica. 1982; 34:316–323. [PubMed: 7152410]

7. Trojan F, Kryspin-Exner K. The decay of Articulation under the influence of alcohol and paraldehyde. Folia Phoniatrica. 1968; 20:217–238. [PubMed: 5671974]

8. Lester, L.; Skousen, R. The Phonology of Drunkenness. In: Bruck, A.; Fox, RA.; LaGaly, MW., editors. Papers from the Parasession on Natural Phonology. Chicago: Chicago Linguistic Society; 1974.

9. Seltzer ML, Vinokur A, Van Rooijen L. A self-administered short Michigan Alcoholism Screening Test (SMAST). J Studies Alcohol. 1975; 36:117–126.

10. MacAndrew C. The differentiation of male alcoholic outpatients from nonalcoholic psychiatric outpatients by means of the MMPI. Quart J Stud Alcohol. 1965; 26:238–246. [PubMed: 14320345]

11. Gough, HG. Manual for the California Psychological Inventory. Palo Alto, CA: Consulting Psychologists Press; 1969.

12. Hoffman H, Loper R, Kammeier M. Identifying future alcoholics with MMPI alcoholism scales. Quart J Stud Alcohol. 1974; 35:490–498. [PubMed: 4154478]

13. Jones JC. Personality correlates and antecedents of drinking patterns in adult males. J Consult Clin Psychol. 1968; 32:2–12. [PubMed: 5649201]

14. Borden, GJ. PhD dissertation. City University; NY: 1971. Some effects of oral anesthesia upon speech: A perceptual and electromyographic analysis.

15. Luce, P.; Carrell, T. Research on Speech Perception: Progress Report No. 7. Speech Research Laboratory, Indiana University; Bloomington, IN: 1981. Creating and editing waveforms using WAVES; p. 287-297.

16. Edwards, AL. Experimental design in psychological research. 2. New York: Holt, Rinehart, and Winston; 1960.

17. Borden GJ, Harris KS, Catena L. Oral feedback II. An electromyographic study of speech under nerve-block anesthesia. J Phonetics. 1973; 1:297–308.

18. Borden GJ, Harris KS, Oliver W. Oral feedback I. Variability of the effect of nerve-block anesthesia upon speech. J Phonetics. 1973; 1:289–295.

19. Wilson JR, Plomin R. Individual differences in sensitivity and tolerance to alcohol. Soc Biol. 1986; 32:162–184. [PubMed: 3837348]

20. Borden, GJ.; Harris, KS. Speech Science Primer: Physiology, Acoustics, and Perception of Speech. Baltimore: Williams & Wilkins; 1980.
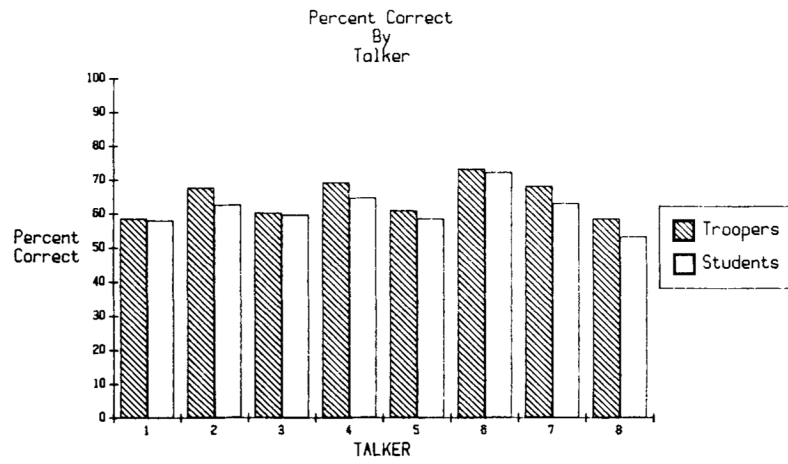
Fig. 1.
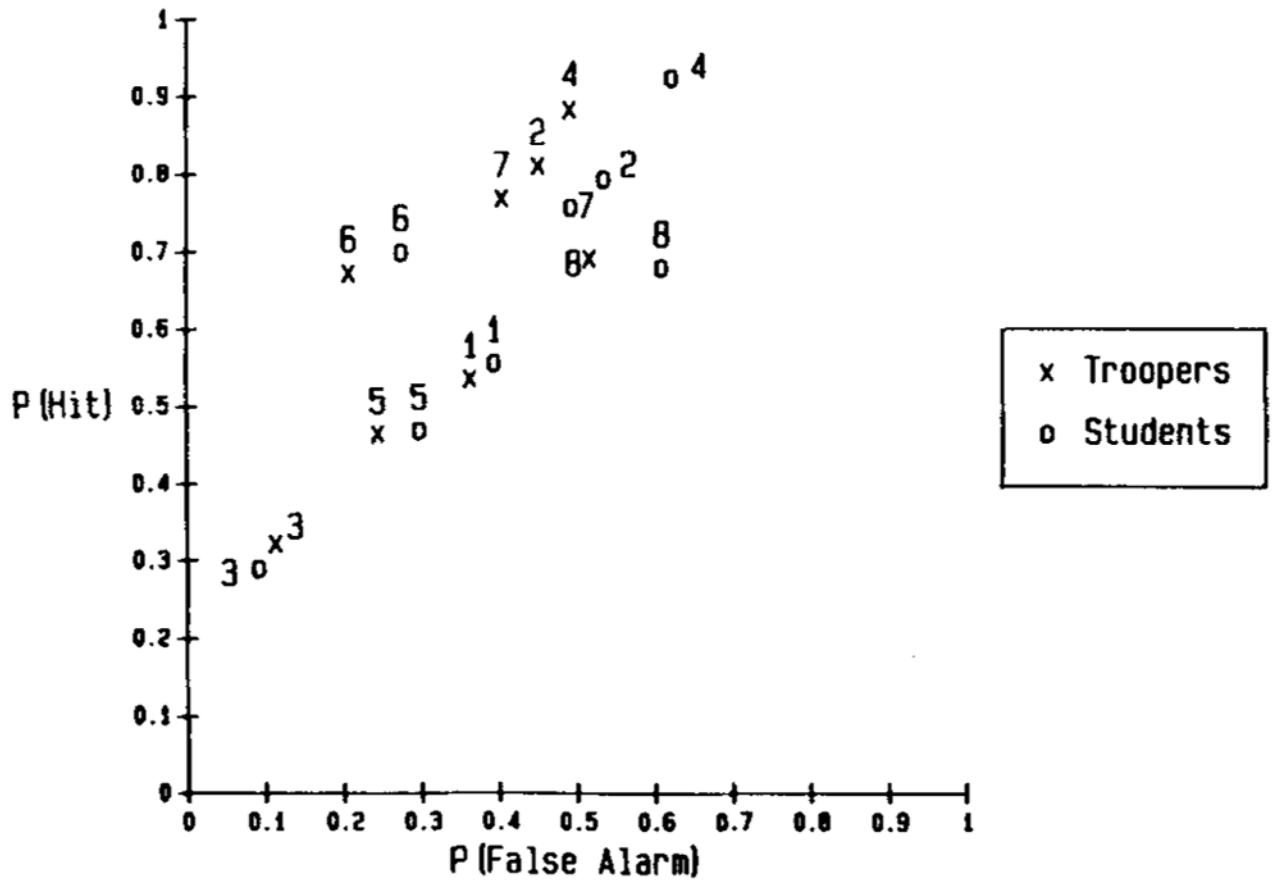Listeners' mean percentage of correct responses by talker (second perceptual experiment).

Fig. 2.
ROC space for individual talkers by listener group (second perceptual experiment).
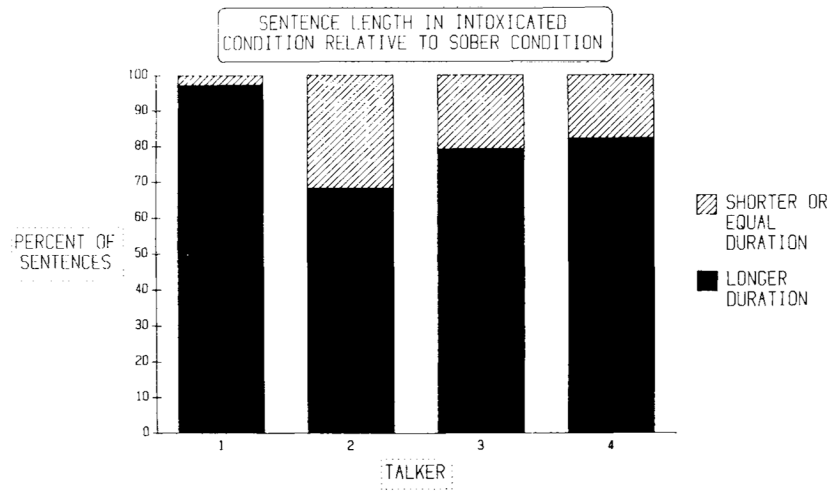
Fig. 3.
Percentage of sentences produced under intoxication which had a longer duration than the same sentences produced by the same talkers while sober.
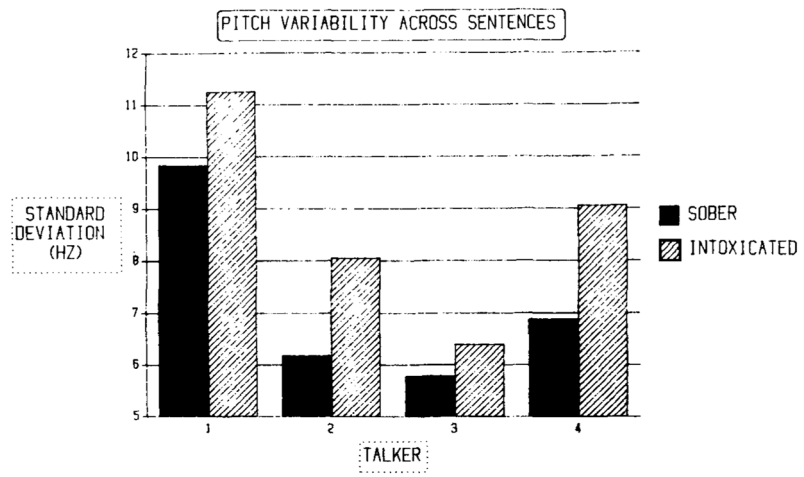
Fig. 4.
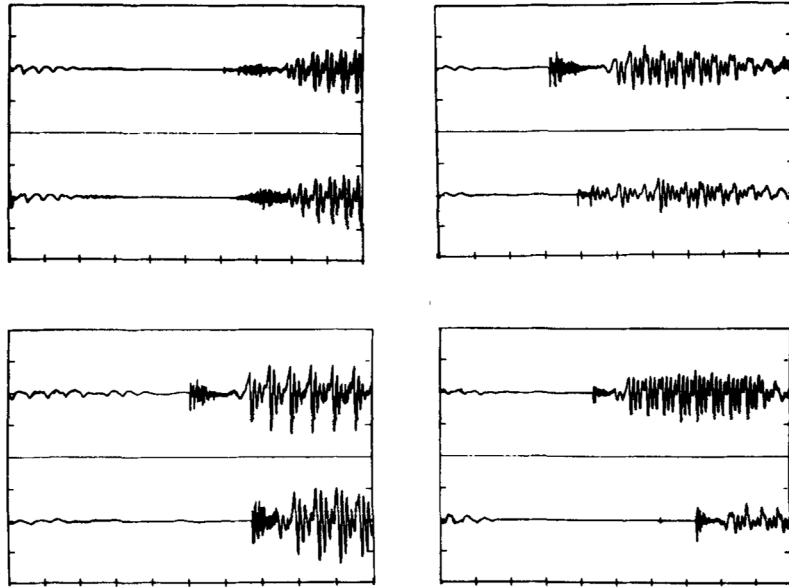Pitch variability across sentences produced by talkers while sober and intoxicated.

Fig. 5.
Waveforms of the /d/ sound in the word "dishes" produced by four talkers. Top trace: sober condition, bottom trace: intoxicated condition, (time between tick-marks, 19.2 msec).
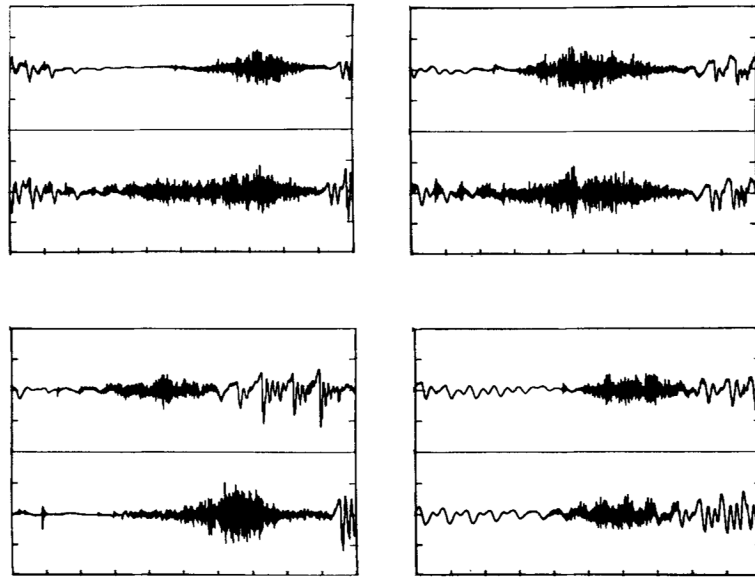
Fig. 6.
Waveforms of the /ʤ/ sound in the word "pajamas" produced by four talkers. Top trace: sober condition, bottom trace: intoxicated condition, (time between tick-marks, 12.8 msec).
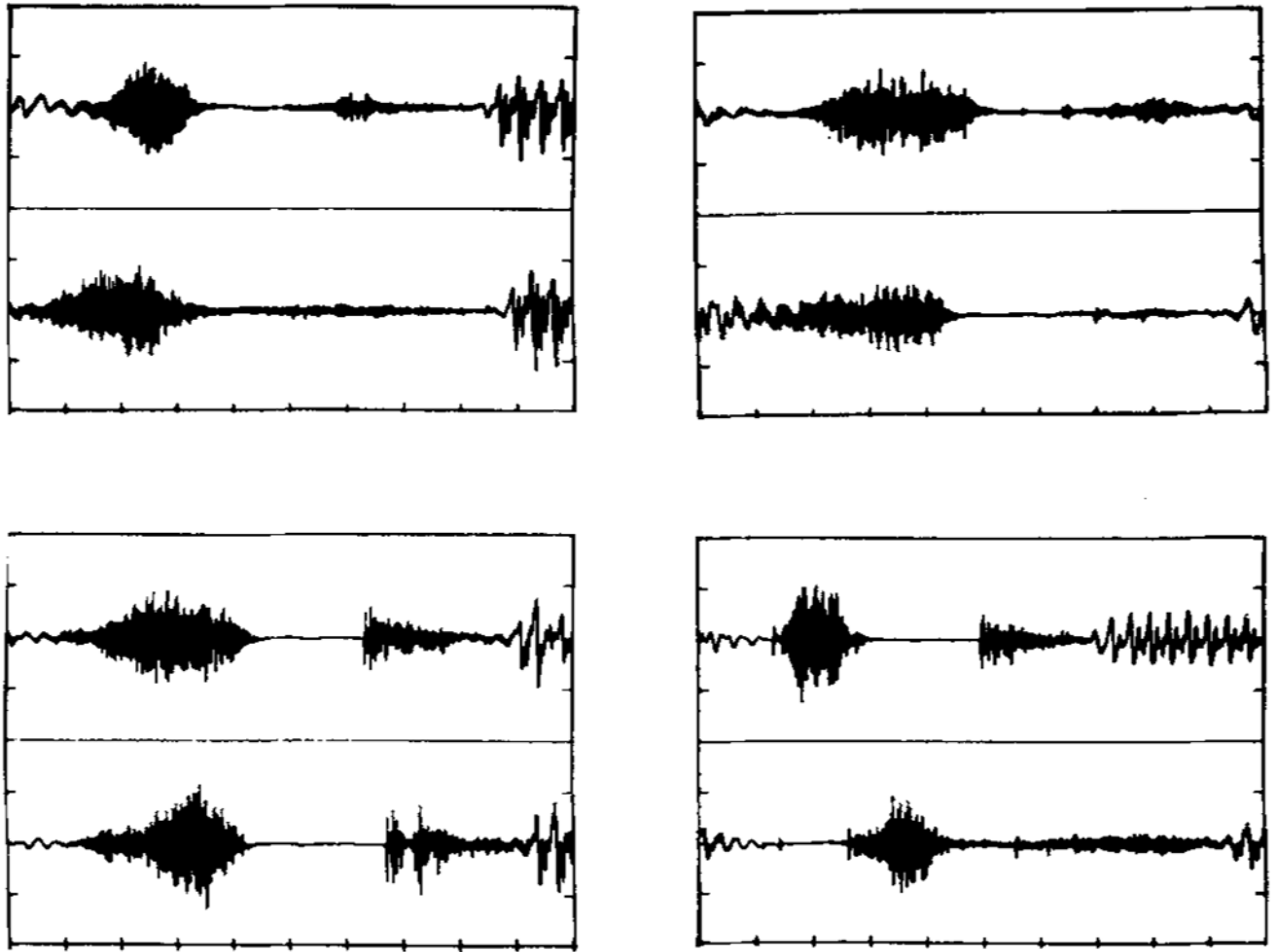
Fig. 7.
Waveforms of the /dʒ/ and /k/ sounds in the end of the word "garbage" and the beginning of the word "cans" produced by four talkers. Top trace: sober condition, bottom trace: intoxicated condition, (time between tick-marks, 19.2 msec for talkers in top panels, 25.6 msec for talkers in bottom panels).

Table 1

Talkers' BACs at the Beginning of the Recording Session

Talker	BAC
1	0.10%
2	0.17%
3	0.15%
4	0.16%
5	0.15%
6	0.13%
7	0.19%
8	0.13%

Table 2

Listeners' Mean Percentage of Correct Responses for the Four Talkers: First Perceptual Experiment

Talker	Percent correct
1	72%
2	70%
3	80%
4	73%

Table 3

Percentage Accuracy Across the Five Confidence Rating Categories for Trooper and Student Listeners: Second Perceptual Experiment

Confidence ratings (1 = least confident; 5 = most confident)	Listener group	
	Troopers	**Students**
1	56.9%	50.0%
2	56.0%	53.6%
3	61.6%	58.7%
4	69.5%	64.8%
5	75.3%	72.5%

Table 4

Phonetic Errors in Key Words Produced by the Four Talkers While Sober and After Alcohol Consumption. I = Intoxicated, S = Sober

| | Talker number | | | |
| | 1 | 2 | 3 | 4 |
Error	I(S)	I(S)	I(S)	I(S)
Vowel lengthening	4(0)	1(0)	8(1)	1(0)
Deletions and partial				
Articulations	2(0)	2(0)	3(0)	6(0)
Deaffrication	1(0)	1(0)	2(0)	2(1)
Consonant lengthening	1(0)	0(0)	1(0)	0(0)
/s/-Distortions	0(0)	2(0)	1(0)	1(0)
Devoicing	0(0)	0(0)	1(0)	0(0)
Velarization	0(0)	0(0)	0(0)	1(0)
Total	8(0)	6(0)	16(1)	11(1)