

Do Voice Recordings Reveal whether a Person Is Intoxicated? A Case Study¹

Keith Johnson, David B. Pisoni, Robert H. Bernacki

Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, Ind., USA

Abstract. In this report we consider the possibility that speech analysis techniques may be used to determine whether an individual was intoxicated at the time that a voice recording was made, and discuss an analysis of the speech produced by the captain of the *Exxon Valdez* recorded at several points around the time of the accident at Prince William Sound, Alaska. A review of previous research on the effects of alcohol and other effects on speech production suggests that it may be possible to attribute a certain, unique pattern of changes in speech to the influence of alcohol. However, the rate of occurrence of this pattern or the reliability of a decision based on observations such as these is not known. Acoustic-phonetic analysis of a small number of tokens of Captain Hazelwood's speech recorded before, during and after the accident revealed a number of changes in speech behavior which are similar to the pattern of changes observed in previous laboratory-based research on the effects of alcohol on speech production. We conclude with a discussion of the limitations in making inferences concerning the state of the speaker upon the basis of phonetic data and then discuss several possible explanations of the pattern of change found in the recordings of Captain Hazelwood.

In this report, we briefly summarize previous research on the effects of alcohol and

other environmental and emotional factors on speech production. We then discuss an analysis of the speech produced by the captain of the *Exxon Valdez* recorded at several times before, during, and after the accident at Prince William Sound. The tapes that we analyzed and information concerning the communications/recording equipment and the times of the recordings were provided to us by the staff of the National Transportation Safety Board (NTSB).

¹ The analyses reported in this paper were carried out in connection with the National Transportation Safety Board (NTSB) investigation of the *Exxon Valdez* accident that occurred on March 24, 1989. This is a revised version of the report which the authors submitted to the NTSB in May 1990. We wish to acknowledge the comments and criticisms of Dr. Malcolm Brenner and our colleagues at Indiana University.

The Problem of Unique Specification

Before discussing Captain Hazelwood's speech, we wish to place the present investigation within a general framework. The question which we are attempting to address in this report is whether it is possible to determine if an individual was intoxicated at a particular time based on acoustic analyses of voice recordings. This question hinges crucially on whether there are properties of speech which occur when a speaker (any speaker) is intoxicated and which do not occur in any other circumstance. We will call this the problem of unique specification.

In the following section, we review several studies which have found that there are a number of acoustic-phonetic characteristics of speech which occur when individuals are intoxicated. This research is an important first step in determining whether speech patterns may uniquely specify alcohol intoxication, but, to our knowledge, there is no published research which directly addresses the problem of unique specification. In spite of this lack of previous research, there are at least two reasons to believe that voice recordings may contain reliable information which uniquely indicates that an individual was intoxicated at the time of the recording. These have to do with the physiological and pharmacological effects of alcohol and the complexity of speech motor control.

Berry and Pentreath [1980] review some of the data having to do with the effects of alcohol on neural membrane permeability and the synthesis and release of neurotransmitter. They note a variety of specific cellular effects and affected sites in the nervous system. Thus, although the effects of alco-

hol at a cellular level in the nervous system are not fully understood, the general functional effects are clear: 'The principal effects of acute dosage of ethyl alcohol are observed in the nervous system, where there is a progressive and simultaneous impairment of function at many levels' [Berry and Pentreath, 1980, p. 43]. Ethanol diffuses easily through cell boundaries [Wallgren and Barry, 1970, p. 36] and results in a biphasic neural response. At low concentrations, nerve cell excitability is increased, while at high concentrations there is a progressive reduction of excitability (p. 254). This reduction in nerve cell excitability leads to behavioral responses to alcohol which (particularly relevant for speech) include decreased motor coordination.

In addition to the neurological effects of alcohol when it reaches the brain through the blood stream, it is likely that the local contact of alcohol with the surfaces of the oral cavity can have some effect on speech production. It is well known that local concentrations of alcohol in the stomach irritate the mucosa and paralyze the muscles of the stomach wall [Wallgren and Barry, 1970, pp. 40, 61]. There is also some evidence suggesting that alcohol applied to the tongue (at least the tongues of cats) can produce a biphasic sensitivity to mechanical stimulation [Hellekant, 1965]. These local effects of alcohol in the mouth and throat may result in effects on speech production which differ from the effects produced by other central nervous system depressants or other factors, although we are aware of no previous research which has attempted to test this specific hypothesis.

Tests of motor coordination (such as walking a straight line or standing on one foot with eyes closed) are commonly used

to indicate whether a person is intoxicated. Speech production is another complex motor activity which requires a high degree of coordination, and it is likely that it may also be affected by alcohol consumption. Two types of motor complexity in speech production can be distinguished. First, speech production requires very precise intergestural coordination. For example, the main difference between /d/ and /t/ in English is the timing of a gesture of the vocal folds relative to a gesture performed by the tip of the tongue. The relative timing of these two gestures (voice onset time) is measured in milliseconds (ms) [Lisker and Abramson, 1964]. The onset of voicing (vocal fold vibration) for /d/ in word-initial position occurs approximately simultaneously with the release of oral stop closure, while the onset of voicing for /t/ occurs 40–60 ms after the release of oral stop closure. Mistiming the two gestures by as little as 20 ms results in a perceptually different consonant. Second, speech involves fine motor control in moving the articulators to the target positions for different speech sounds. For example, the fricative /s/ is produced by pressing the sides of the tongue against the upper molars and depressing the center of the tongue, creating a narrow groove with the tip of the tongue. The articulatory difference between /s/ and /ʃ/ is very subtle even though the acoustic difference is quite large [in Stevens', 1972, terms, a 'quantal' effect]. The location of the tongue relative to the front teeth and the length of the constriction at the roof of the mouth (the tongue groove) distinguish these two sounds in speech production [Subtelný et al., 1972]. If the tongue tip is kept close to the front teeth and the constriction at the roof of the mouth is relatively short (2.5 cm), an /s/ is produced.

However, if the constriction is slightly longer or wider, or the tongue tip is held a little further back in the mouth, the resulting sound is more like /ʃ/. These observations suggest that small variations in speech timing or motor control can have acoustically reliable consequences for speech production.

The effects of alcohol on the central nervous system and the local effects of alcohol on the muscles and proprioceptors of the vocal apparatus, coupled with the inherent complexity of speech production, suggest that there may be aspects of speech production which are uniquely altered by alcohol intoxication.

Previous Findings on Alcohol-Impaired Speech

This section is a brief review of previous research on the effects of alcohol on speech production. For more complete reviews of the literature see Pisoni et al. [1986], Klingholz et al. [1988], and Pisoni and Martin [1989]. The effects of alcohol on speech production that have been observed in controlled laboratory studies can be divided into three types: gross effects, segmental effects and suprasegmental effects. Examples of each of these effects are listed in table 1.

Gross effects involve word level alterations in speech production. These effects are very noticeable when intoxicated subjects are instructed to read a passage. Subjects may revise, omit or interject words [Sobell and Sobell, 1972; Sobell et al., 1982]. It has been assumed that this class of speech errors reflects modifications in speech planning. As neural function is depressed by alcohol, the speaker's ability to control the ar-

Table 1. Summary of previous research on the effects of alcohol on speech production

Gross effects	word/phrase/syllable interjections ¹ word omissions ^{1,4} word revisions ¹ broken suffixes ¹
Segmental effects	misarticulation of /r/ and /l/ ^{4,5} /s/ becomes /ʃ/ ^{3,4} final devoicing (e.g. /iz/ → /is/) ^{3,5} deaffrication (e.g. 'church' → 'shursh') ³⁻⁵
Suprasegmental effects	reduced speaking rate ^{1,2,3,5} decreased amplitude ² increase of unvoiced-to-voiced ratio ^{3,5,6} decreased spectral tilt ⁶ mean change in pitch range (talker-dependent) ⁴⁻⁷ increase in pitch variability ^{5,6}

¹Sobell and Sobell [1972]: 16 alcoholics, 5–10 ounces, 86 proof alcohol.

²Sobell et al. [1982]: 16 talkers, 0.05 < BAL < 0.1.

³Lester and Skousen [1974]: Number of talkers not mentioned, 86 proof straight bourbon, one ounce/20 min up to 14 ounces.

⁴Trojan and Kryspin-Exner [1968]: 3 talkers, 1–1.38 liters of heavy Austrian wine (13% alcohol).

⁵Pisoni et al. [1986] and Pisoni and Martin [1989]: 5 talkers, 0.1 < BAL < 0.17.

⁶Klingholz et al. [1988]: 16 talkers, 0.067 < BAL < 0.16.

⁷Dunker and Schlosshauer [1964]: 1 talker, 'consuming alcoholic beverages liberally' and shouting.

articulators is impaired, which in turn may affect the planning stage in speech production. Thus, word level alterations occur when the subject is required to read a passage aloud. In spontaneous speech, however, it is much harder to decide what should count as a gross error because the

speaker's intended utterance is not known. Therefore, gross effects are less valuable for the evaluation of spontaneous speech and diagnosis of any impairment due to alcohol.

Segmental effects involve the misarticulation of specific speech sounds. The segmental effects which have been most often reported are: misarticulation of /r/ and /l/, misproduction of /s/ (more like /ʃ/), final devoicing of obstruents, and deaffrication. Examples of the last two effects are given in table 1. Obstruent devoicing involves a problem of timing and glottal control similar to the example of /d/ and /t/ given in the previous section. The other segmental effects involve the control of the tip of the tongue. Lester and Skousen [1974] found that segmental effects such as these did not appear until subjects had consumed about 10 ounces of 86 proof straight bourbon over a period of about 3.5 h.

Phonetic theory makes some predictions about the modifications of speech articulation after alcohol consumption. These predictions derive from the study of articulatory ease [see for example Lindblom, 1983], which suggests that not all speech sounds are equally easy to produce. Evidence of this comes from studies of the development of speech in children [de Villiers and de Villiers, 1978], the patterns of historical language change [Anttila, 1972], and patterns of language dissolution in aphasia [Jakobson, 1941], as well as model studies of articulation [Lindblom, 1983]. Most of the segmental effects observed in speech produced while talkers were intoxicated have analogs in these data. For instance, it is common for children to misarticulate /r/ and /l/ as in the production 'train' as /twen/. Also, final devoicing and deaffrication are very common in child speech and in historical lan-

guage development. The substitution of /f/ for /s/, however, is not typically found in child speech, and /s/ is more common than /f/ in the languages of the world. Therefore, this segmental effect, rather than being the result of a general loss of motor coordination (as is most likely the case for the other segmental effects), seems to have a different cause. The change of /s/ to /f/ may be related to loss of responsiveness of the surface muscles of the tongue or a loss of proprioceptive feedback from the tongue after direct contact with ethanol during consumption. If this proposed account of the /s/-to-/f/ change is correct, the mispronunciation of /s/ may be a unique phonetic correlate of intoxication.

Suprasegmental effects are perhaps more perceptually salient than segmental effects. Trojan and Kryspin-Exner [1968] reported an increase in fundamental frequency (F_0). Pisoni and Martin [1989] found that F_0 decreased for some, but not all subjects. Klingholz et al. [1988] also found a tendency for decreased F_0 . F_0 is also more variable in speech produced while intoxicated when compared to a control condition [Pisoni and Martin, 1989; Klingholz et al., 1988]. As will be noted below, there are two possible sources for this increased variability. First, intoxicated speech may involve more extreme intonations than normal speech. Second, intoxicated speech may involve increased vocal jitter (period-to-period variation). Both of these effects will produce an increase in the variance of F_0 , all other things being equal. Klingholz et al. [1988] also found that the speech harmonics-to-noise ratio decreased after alcohol intoxication. This measure reflects a change in the mode of vocal fold vibration indicative of increased breathiness after alcohol

intoxication. They also found a change in the long-term average spectrum in intoxicated speech. There was an increase in high-frequency energy, which may reflect, in addition to an increase in the harmonics-to-noise ratio, an increase in the unvoiced/voiced ratio (the ratio of the number of frames classified as voiceless and the number of frames classified as voiced) after alcohol consumption [as reported by Pisoni and Martin, 1989]. Finally, all previous studies which included relevant measures have found that speaking rate decreases (segmental durations increase) after alcohol consumption.

The effects on speaking rate and F_0 can be related to the general physiological effects of alcohol in the following ways. The reduction in speaking rate may be the result of an attempt to compensate for the loss of motor coordination which accompanies intoxication. The effect of alcohol on F_0 seems to have an origin in the interaction of alcohol with the tissue of the vocal folds. Klingholz et al. [1988] suggest that the effect of alcohol on F_0 may be the result of irritation and swelling of the mucous membranes of the vocal folds and desensitization of the proprioceptors of the vocal folds. They cite evidence from Dunker and Schlosshauer [1964] which indicates that vocal fold vibration after alcohol consumption (like vocal fold vibration for people with hoarse voices) is more variable and lower in pitch. Klingholz et al. [1988] posited a connection between vocal fold edema due to mechanical stress [shouting or speaking for an extended time] and edema due to alcohol consumption. This explanation may also account for the increase in the unvoiced/voiced ratio in intoxicated speech. The fact that Trojan and Kryspin-Exner [1968]

found an increase in F_0 while most other studies have found a decrease in F_0 may reflect a biphasic response to alcohol (as noted earlier for the neurological effects of alcohol). Trojan and Kryspin-Exner [1968] did not measure blood alcohol level and were generally vague about their methods, so they may have been reporting an effect which occurred at a lower level of intoxication than the effects reported in the other studies.

Other Effects on Speech Production

In this section, we briefly review some of the previous research on environmental and emotional effects on speech production and compare these effects with the effects of alcohol on speech production. Table 2 is a summary of some previous research addressing environmental and emotional effects on speech production. As indicated in this table, most researchers who have investigated the effects of these factors on speech production have focussed on suprasegmental phenomena. Only occasionally have segmental phenomena, other than vowel formant measures, been investigated. This research focus reflects a practical concern for the design of automatic speech recognition devices for use in a variety of circumstances, where suprasegmental changes and some types of segmental changes could be detrimental to the performance of recognition systems. Therefore, the data base we are reviewing here is not entirely comparable to that collected in the study of the effects of alcohol on speech.

Hansen [1988] and Summers et al. [1988] studied the effects of *noise* on speech production (the Lombard effect). These studies

found that speech produced with a high level of noise at the ears had increased F_0 and duration, and reduced spectral tilt. The spectral-tilt measure indicates that there was a relative increase of high-frequency glottal energy in the Lombard condition. Surprisingly, Hansen [1988] found no change in amplitude. The Summers et al. [1988] result (increased amplitude in the noise condition) is in better agreement with earlier research on the Lombard effect. Finally, the studies indicate some individual variability in the effect of noise on vowel formant values.

Hansen [1988] measured the tilt of the glottal spectrum (after inverse filtering) while the other authors listed in table 2, who reported spectral tilt changes, measured changes in the spectral tilt of the unfiltered speech signal. There is general agreement between studies using the two measures, although note that valid tilt comparisons using the simpler method require careful control of the phonetic content (particularly vowel qualities) of the tokens being compared.

Moore and Bond [1987] studied the effects of *acceleration* (speakers spun in a centrifuge) and *vibration* (speakers seated on a vibrating table) on speech produced by 2 subjects. The two situations resulted in comparable effects on F_0 , intensity, and vowel formants. F_0 increased relative to that found for the same subjects in benign environments, vocal intensity was unchanged, and vowels were less distinctive (more like /ə/). There was individual variability in the effect of acceleration on segmental duration, while speaking rate increased (reduced segmental durations) in the vibration condition. The small number of subjects in these studies is problematic, but these are the only available data on these environmental effects.

Table 2. Summary of some recent research on environmental and emotional effects on speech production

Condition	F ₀	SD F ₀	Jitter	Tilt	Duration	Intensity	Formants
Noise ¹	↓	↓		↓	↓	NC	F ₁ ↓
Noise ²	↓			↓	↓	↓	F ₁ ↑
Acceleration ³	↓				↓	NC	centralized
Vibration ³	↓				↓	NC	centralized
Workload ⁴	↓	↓		↓	NC	↑	F ₁ and F ₂ ↓
Workload ⁴	↑	↓		↓	↓	↑	NC
Workload ⁵	↓	↑			↓	↓	
Stress ⁶	↑		↓		↓	↓	
Stress ⁷	↑		↓				
Stress ⁸	↓				↓	↑	
Perceived stress ⁸	↓				↓	↓	
Fear ⁹	↓	↓	↓	NC	↓		NC
Fear ¹	↓	↓		↓	NC	↓	F ₁ and F ₂ ↓
Anger ⁹	↓	↓	NC	↓	NC		F ₁ ↓
Anger ¹	↓	↓		↓	↓	↓	F ₁ ↑
Sorrow ⁹	↓	↓	↓	↓	↓		NC
Depressed ¹					↓		centralized
Intoxicated ¹⁰	↓	↓		↓	↓	↓	

↓ = Reliable *increase* for all subjects; ↑ = increase for some, but not all subjects; ↓ = reliable *decrease* for all subjects;

↓ = decrease for some, but not all subjects; ↓ = some subjects showed a reliable increase, while some a reliable decrease; NC = no change.

¹Hansen [1988] (8 talkers).

²Summers et al. [1988], see also Pisoni et al. [1985] (2 talkers).

³Moore and Bond [1987] (2 talkers).

⁴Summers et al. [1989] (5 talkers).

⁵Griffin and Williams [1987] (20 talkers).

⁶Brenner and Shipp [1988] (17 talkers).

⁷Brenner et al. [1985] (7 talkers).

⁸Streeter et al. [1983] (2 talkers).

⁹Williams and Stevens [1972], see also Williams and Stevens [1981] (3 talkers).

¹⁰See table 1.

A large number of studies have employed *workload* tasks to simulate environments with high cognitive demands such as airplane cockpits. These studies have generally found that speech produced while performing a cognitively demanding task has higher F₀, decreased spectral tilt and in-

creased intensity [Hansen, 1988; Summers et al., 1989, Griffin and Williams, 1987]. Data on the variability of F₀ (SD F₀) are mixed. This reflects a problem in the use of this measure due to the fact that F₀ variability can be affected in two very different ways. Variability will be reduced if the F₀

contours of utterances are more monotonic in the workload condition [as suggested by Summers et al., 1989] or if there is less period-to-period variation in the vibratory cycle of the vocal folds [as suggested by Brenner et al., 1985]. On the other hand, F_0 variability could be increased if utterances in the workload task had more extreme fluctuations in their F_0 contours even if vocal fold jitter (period-to-period variation of F_0) were reduced. Williams and Stevens [1972] provide a good example of the conceptual distinctions which need to be maintained in this area, although they did not have digital signal processing techniques at their disposal. They reported both changes in F_0 range and (inferences about) changes in F_0 jitter. In the absence of this distinction in some of the research on the effects of cognitive workload, it is impossible to determine whether the reported differences in F_0 variability in speech under workload reflect real individual differences or merely differences in data collection techniques. Table 2 also indicates some differences across studies in the effects of workload on segmental duration, although it is interesting that the study on the effects of workload which employed the greatest number of subjects [Griffin and Williams, 1987] reported a consistent decrease in duration. Finally, there is also some discrepancy concerning the effects of workload on vowel formant frequencies.

The term *psychological stress* has been used to describe situations ranging from lying to being in a fatal airplane crash. Scherer [1981] outlined some predictions for speech production in stressful situations based on the general physiological response to stress (similar to the discussion above of physiological predictions for the effects of alcohol) and then concluded that 'virtually

all of the studies in this field have found very strong individual differences in terms of the number and kind of vocal parameters that seem to accompany stress' (p. 179). He focussed on two problems in the literature, (1) the likelihood that subjects in laboratory studies of stress were differentially stressed, and (2) the fact that 'subjects may differ in terms of the degree of control they can exert as far as their vocal production under emotional arousal is concerned' (p. 180). (Both of these problems have analogs in studies of the effects of alcohol on speech. Although, it is possible to objectively measure the subjects' blood alcohol level (BAL), not all previous research on the effects of alcohol on speech production have reported BALs. Also, subjects may differ in the degree of articulatory control they can exert while intoxicated.) In spite of these problems, some general trends emerge from studies of stress in laboratory and real-life emergency situations. These are indicated in table 2 and include an increase in F_0 , an increase in intensity, and a decrease in F_0 jitter. Brenner et al. [1985] examined F_0 jitter in situations of high stress by analyzing voice recordings of pilots involved in aircraft crashes. They found that speech in stressful situations had increased F_0 , and decreased F_0 jitter. In a related laboratory study, Brenner et al. [1985] also found that the activity of the cricothyroid muscle, which is the primary muscle of the larynx involved in controlling F_0 , increased as stress increased. This provides an explanation of both the increased F_0 and decreased F_0 jitter found in the other studies if we assume that increased muscular tension in the glottis allows for less variability in the pattern of vocal fold vibration while causing a higher rate of vibration.

Streeter et al. [1983] reported a case of individual variability in the vocal effects of stress. They examined a recorded telephone conversation between a system operator and chief system operator for Consolidated Edison during the New York City blackout, July 1977. One talker had increased F_0 , duration, and amplitude as the situation developed (and presumably stress increased), while the other showed a different pattern (decreased F_0 and duration, and no change in amplitude). This study supports Scherer's [1981] point about individual differences in response to stressful situations and suggests that there may be no consistent phonetic pattern across subjects for any but the most extremely stressful, life-threatening situations. Interestingly, though, Streeter et al. [1983] found that naive listeners used phonetic cues consistently in making judgements about the degree of stress being experienced by the talker. Listeners judged utterances with higher F_0 , higher amplitude and longer segment durations as more stressed even though, for 1 speaker, these judgements were not correlated with the degree of experienced stress. The speech parameters which were found in this study to be correlated with *perceived stress* are listed in table 2. Streeter et al. [1983] concluded that listeners have stereotyped expectations for vocal responses to stress, which evidently are accurate for the most extreme levels of stress, but speakers who are actually experiencing some less than maximal degree of stress do not always fit the perceptual stereotype.

Table 2 also presents a summary of several studies on the effects of emotional state (fear, anger, and sorrow) on speech production. The study of the effects of emotion on speech production involves method-

ological problems that are not involved in the study of environmental effects on speech, where it is possible for the experimenter to create conditions which can be carefully controlled and described. In order to study the effects of emotion on speech production, however, it is necessary to rely on subjective measures of the emotional (mental) state of the speaker or have speakers simulate various emotions. In spite of these methodological difficulties, we are including this summary of previous research in an attempt to present a complete review of the factors that may affect speech production.

Williams and Stevens [1972, 1981] hired 3 actors to perform short plays in which the characters displayed various emotions. Their data are summarized in table 2 and compared with some recent data from Hansen [1988], who studied the effect of *fear* by having his subjects read a prepared wordlist as they were descending steep drops on a roller coaster. There is good agreement between these two studies concerning the effects of fear on F_0 . Both found that F_0 increased and that F_0 variability increased. Williams and Stevens also suggested that, in addition to increased F_0 range, F_0 jitter increased. Whereas Williams and Stevens reported no change in spectral tilt, Hansen found that the glottal spectrum was flatter in the fear condition. The more sophisticated signal-processing techniques employed by Hansen may have allowed him to detect a small change not seen by Williams and Stevens. The two studies also found different effects on segmental duration. Hansen found no change, while Williams and Stevens found an increase in word duration of about 30 ms. This seems to reflect a real difference, and again may be a result

of methodological differences. Hansen reported that intensity increased in the fear condition. This effect is consistent with findings for psychological stress and increased workload and seems to reflect a change in arousal [Scherer, 1981]. Finally, Hansen found changes in the first two vowel formants which were not found by Williams and Stevens.

Hansen [1988] and Williams and Stevens [1972] also studied the effects of *anger* on speech production. Here the two studies had similar methodologies and very similar results. They both found that F_0 , F_0 variability and F_1 increased, and that spectral tilt decreased. Williams and Stevens found no changes in F_0 jitter, although they were using a somewhat crude measure (fluctuation in narrow-band spectrograms). Hansen found an increase in intensity. The only discrepancy between the two studies has to do with the effect of anger on speaking rate. Where Williams and Stevens found no reliable change, Hansen found that speaking rate decreased (increased segmental durations) in the anger condition. Notice the similarities between the effects of anger and the effects of workload and fear.

The final emotion listed in table 2 is *sorrow*. Again, the data listed in the table are from William and Stevens [1972] and Hansen [1988]. The data reported by Hansen are based on a small number of observations. These data are included in the table because they come from a real-life situation (recordings made during counselling sessions in a psychiatrist's office) and as such offer some degree of validation of the observations of Williams and Stevens [1972]. Speech produced by actors portraying sorrow was characterized by decreased F_0 , decreased F_0 range, but increased F_0 jitter.

Williams and Stevens also found that spectral tilt increased in the sorrow condition (i.e. that there was a reduction of high-frequency energy). Both Hansen and Williams and Stevens found an increase in segmental durations, but they found different effects on vowel formants. Williams and Stevens found no change in vowel formants while Hansen suggested (based on very few measurements) that vowels were more centralized in the depressed condition.

We have also included in table 2 a summary of the suprasegmental effects found in the studies of alcohol and speech which were listed in table 1. There are no situations or emotions listed in table 2 which have exactly the same pattern of effects found in the studies of alcohol and speech, and so, given adequate measures of these acoustic correlates, it is in principle possible to classify the changes observed across two or more samples of speech as more like the pattern found for intoxicated speech than, for instance, speech produced in noise. It is not possible, however, to give any kind of confidence rating to such a classification, because there is not enough published data on individual differences which would allow the calculation of hit rates and false alarm rates for classifications based on these measures (this is true of the other effects shown in table 1 also).

Another problem with classifying speech samples is that there are some possible physiological effects on speech production which have not been previously studied. The effect of fatigue on speech production has not been examined in any controlled study. Also, we lack any data on speech production just after the speaker has been awakened. Our subjective impression is that speech produced in these circumstances

may involve changes in vocal fold activity (extremely low F_0 or pulse register phonation), decreased speaking rate, and perhaps some effects related to dehydration of the mucous membranes in the mouth (e.g. the pronunciation of /s/), which may be similar to the effects seen after alcohol consumption. However, the relevant controlled laboratory studies have not been done, also there are no data on more complex situations involving combinations of effects. For instance, no one has studied the combined effects of fatigue and stress on speech production.

The Speech of Captain Hazelwood

We analyzed five different samples of speech provided to us by NTSB. Also, we examined a small number of utterances from Captain Hazelwood's televised interview with Connie Chung which was broadcast on March 31, 1990. We will refer to the speech samples according to the times at which they were recorded: 33 h before the accident (-33), 1 h before the accident (-1), immediately after the accident (0), 1 h after the accident (+1), 9 h after the accident (+9), and televised interview (CC). We will discuss gross errors, segmental changes, and suprasegmental changes.

It is important to note here that the recording made 33 h before the accident has a different history than the other recordings. All of the NTSB recordings were initially recorded using the same Coast Guard equipment, but this sample was then rerecorded onto a handheld cassette recorder before the original tape was mistakenly erased. The recording which we analyzed was produced by playing back the cassette tape using the same cassette recorder which had been used to record the sample. We investigated the possibility that the recording

Table 3. Summary of phenomena found in the analysis of the NTSB tape (numbers in parentheses indicate the time of recording)

Gross effects	revisions (-1) Exxon Ba, uh Exxon Valdez (-1) departed disembarked (-1) I, we'll (-1) columbia gla, columbia bay
Segmental effects	misarticulation of /r/ and /l/ (0) northerly, little, drizzle, visibility (/s/ becomes /ʃ/ (fig. 3) final devoicing (e.g. /z/ → /s/) (-1,0,+1) Valdez → Valdes
Suprasegmental effects	reduced speaking rate (fig. 4, 5) mean change in pitch range (talker-dependent, fig. 6) increased F_0 jitter (fig. 6)

was corrupted by analyzing an unidentified background sound which seemed to be present in both the -33 sample and in the -1 sample. In the -33 recording, the sound had a higher average F_0 (480 Hz, $n=4$ versus 472 Hz, $n=10$) and a greater F_0 range (438–588 Hz versus 456–481 Hz) as compared with the -1 recording. The variability of the F_0 in the -1 recording suggests that the sound was not constant in frequency and, thus, is not an adequate benchmark for determining the validity of the -33 recording. However, even if the -33 recording is corrupted by tape speed fluctuations of the magnitude indicated by these measurements (-9 to +22%), this degree of difference is not enough to account for the changes in speech production we report below.

Gross Errors

Several of the speech errors in the NTSB tapes may be classified as gross phonetic or lexical errors. These are listed in table 3. Note, however, that such phenomena are not uncommon in spontaneous speech regardless of alcohol consumption. What is

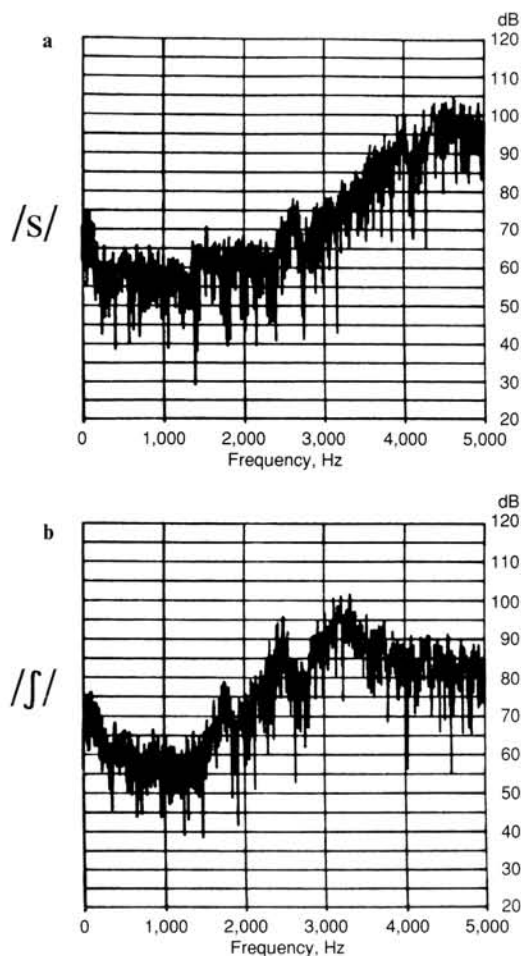


Fig. 1. Power spectra of /s/ (a) and /ʃ/ (b) produced by K. J. in a quiet recording booth with recording equipment responsive up to 5,000 Hz.

needed in order to evaluate the condition of the speaker is a large amount of speech in which it is possible to compare the rate of occurrence of such errors across speech samples. Also, since the talker was not reading a prepared text, it is a matter of subjective judgement to say that something is or is not an error. To attempt to control for this

problem, we are only reporting cases in which the speaker corrected himself. As indicated in table 3, the only examples of gross speech effects that we found in the NTSB tapes occurred in the recording made 1 h before the accident.

Segmental Phenomena

Also in table 3, we have listed some examples of segmental errors. The problem with these data is that the recordings are noisy. Identifying most of the examples listed in the table required repeated listening and phonetic transcription (the exception is the /s/-/ʃ/ example). The amount of noise on the tape increases the probability that the transcriptions are inaccurate. Therefore, we performed acoustic analyses of several productions of /s/.

Figure 1 shows power spectra of /s/ and /ʃ/ produced by the first author (K.J.). As illustrated, /s/ is characterized by a peak of energy in the range from 4,000 to 5,000 Hz, while /ʃ/ has a lower frequency peak (in the range from 3,000 to 4,000 Hz) and a lower amplitude peak of energy in the range from 2,000 to 3,000 Hz. The spectra in figure 1 illustrate what the power spectra of /s/ and /ʃ/ look like in recordings which have a high signal-to-noise ratio and frequency information up to 5,000 Hz [Borden and Harris, 1984, p. 189].

Figure 2 shows power spectra of the /ʃ/ productions of *shout* and *she's* (and spectra of background noise near the fricative) as spoken by Captain Hazelwood in the recording made 33 h before the accident. The lower amplitude peak between 2,000 and 3,000 Hz, illustrated in figure 1, is present in the spectra in figure 2, but the higher frequency information which would normally serve as the most reliable information dis-

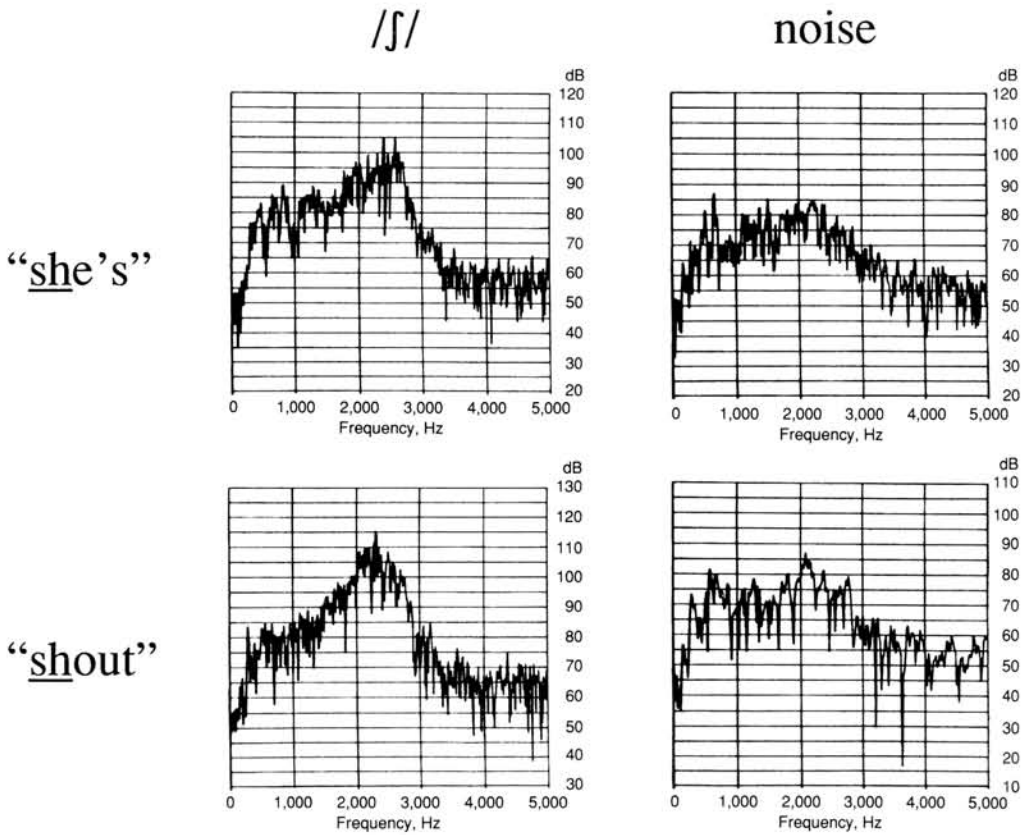


Fig. 2. Power spectra of /ʃ/ produced by Captain Hazelwood in the words *she's* and *shout* recorded 33 h before the accident. Each spectrum is paired with a spectrum of the background noise from a nearby open-mike pause.

tinguishing /s/ and /ʃ/ is not present in these spectra because the radio transmission equipment was band-limited at 3,000 Hz. (Energy above 3,000 Hz was attenuated at approximately 50 dB per octave with a noise floor 50 dB below maximum signal level.) In making these comparisons, we had to be concerned also about the spectral shape of the background noise in the NTSB recordings. The spectra in figure 1 were calculated from recordings made in a quiet recording booth, while the NTSB record-

ings have background noise which may be confused with fricative noise. Therefore, paired with each fricative spectrum from the recordings, we also show a spectrum of nearby background noise as a baseline against which the fricative spectrum can be compared.

Figure 3 shows power spectra of the /s/ of *sea* (or *see*) from the five different recordings paired with spectra of background noise from the same recording. The noise spectra were taken from nearby, open-mike

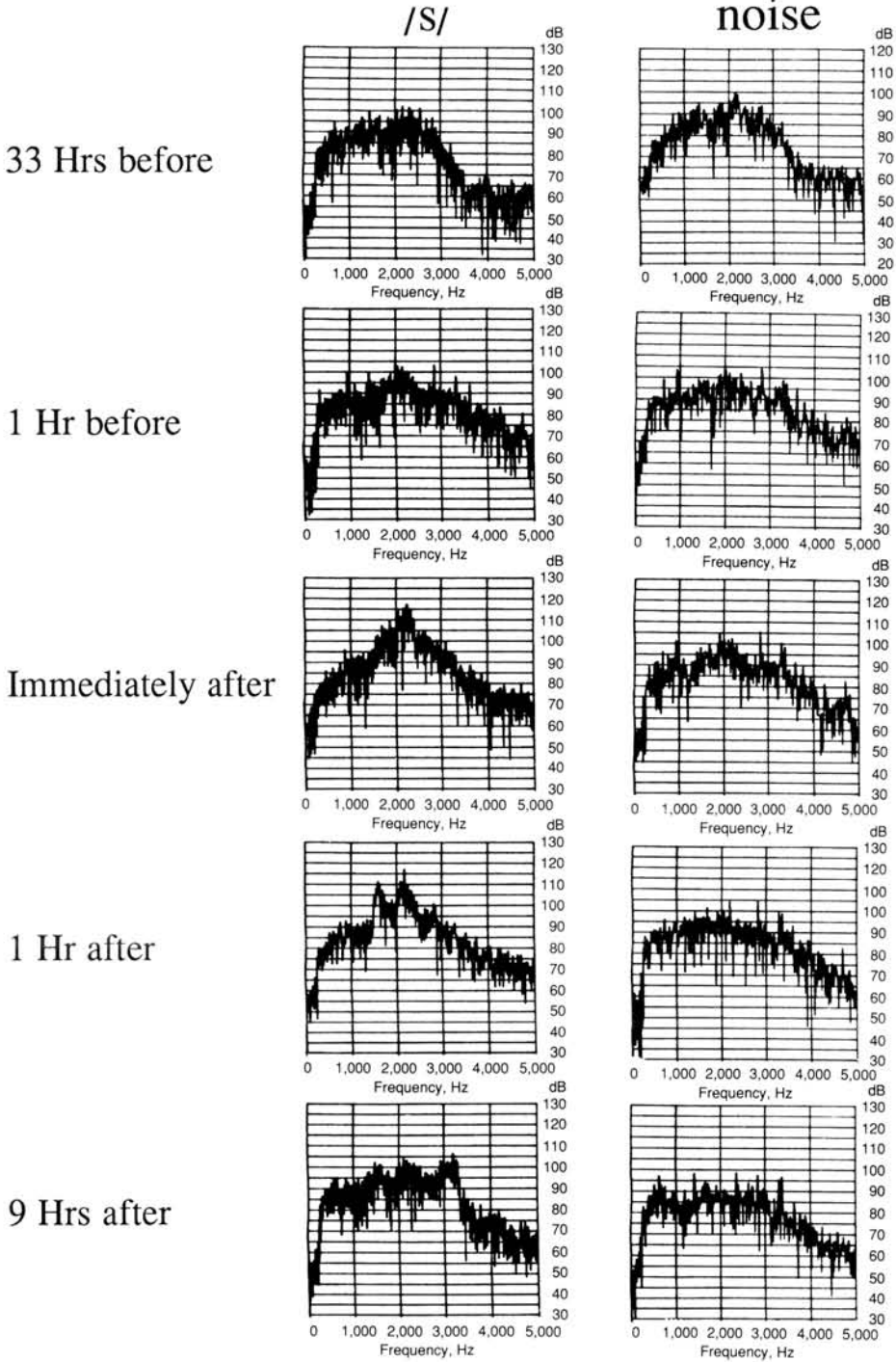


Fig. 3. Power spectra of */s/* paired with spectra of nearby open-mike pauses from each of the NTSB recordings.

background noise. On average the noise segments were 1.3 s from the /s/ segments.

We estimate that the signal-to-noise ratio in these samples ranges from 5 to 10 dB. This estimate of signal-to-noise ratio was taken from measurements of background noise during stop closures because the transmission equipment had an automatic gain control making amplitude measures from pauses inappropriate. Note also that this means that the amplitudes of the background noise spectra in figures 2 and 3 do not accurately reflect the amplitude of background noise in the fricative spectra.

The /s/ spectrum from the earliest recording (33 h before the accident) has the same basic shape that the background noise has, suggesting that the /s/ is buried beneath the noise, or more accurately, that the main spectral energy for /s/ is not within the frequency range of the transmission system. The same is true for the /s/ of *sea* recorded 1 h before the accident. The spectra of /s/ from the recordings made immediately after the accident and 1 h after the accident have peaks of energy (relative to the background noise) in the region from 2,000 to 3,000 Hz. Finally, the spectrum of /s/ recorded 9 h after the accident does not have a peak of energy in the region from 2,000 to 3,000 Hz. We interpret the peaks in the /s/ spectra from samples recorded immediately after the accident and 1 h after the accident as evidence for a segmental change from /s/ to /ʃ/. There is no evidence in these spectra, nor in the other /s/ spectra which we examined, for this segmental change between the earliest recording and the one made 1 h before the accident. These spectral changes reflect a change in the articulation of /s/ similar to that which has been observed in earlier studies of the effects of alcohol on speech production [Lester and Skousen, 1974; Trojan and Kryspin-Exner, 1968]. There is no

data in the literature to indicate whether a change of this sort is to be expected as a result of other factors.

Suprasegmental Properties

Finally, we examined the suprasegmental properties of the speech samples. Because the communication equipment had an automatic gain control and the distance between the microphone and the speaker's lips was (presumably) variable, it is inappropriate to compare measurements of speech amplitude or long-term average spectra. Therefore, we focussed our attention on speaking rate and voice F_0 . We took care to control for discourse position and the position of words within sentences because these factors can affect the suprasegmental properties of speech [Lehiste, 1970; Klatt, 1976]. We analyzed two phrases, *Exxon Valdez* and *thirteen and sixteen*, because these phrases were repeated several times during the recordings and occupied comparable positions in discourse and sentence contexts across the different recordings. Thus, these phrases provide a measure of experimental control which is needed in making valid suprasegmental comparisons across speech samples.

Figure 4 shows durations of the segments in *Exxon Valdez* from each of the recordings. Each bar in this figure is the average of two occurrences of the phrase. As indicated in figure 4a, it took longer to say the phrase in the samples recorded near the time of the accident. Figure 4b (which is another plot of the same data) shows that this effect was more pronounced for the vowels and the /v/ of *Valdez*. If we take this as an index of speaking rate, it is reasonable to conclude from these measurements that the captain was speaking more slowly in the

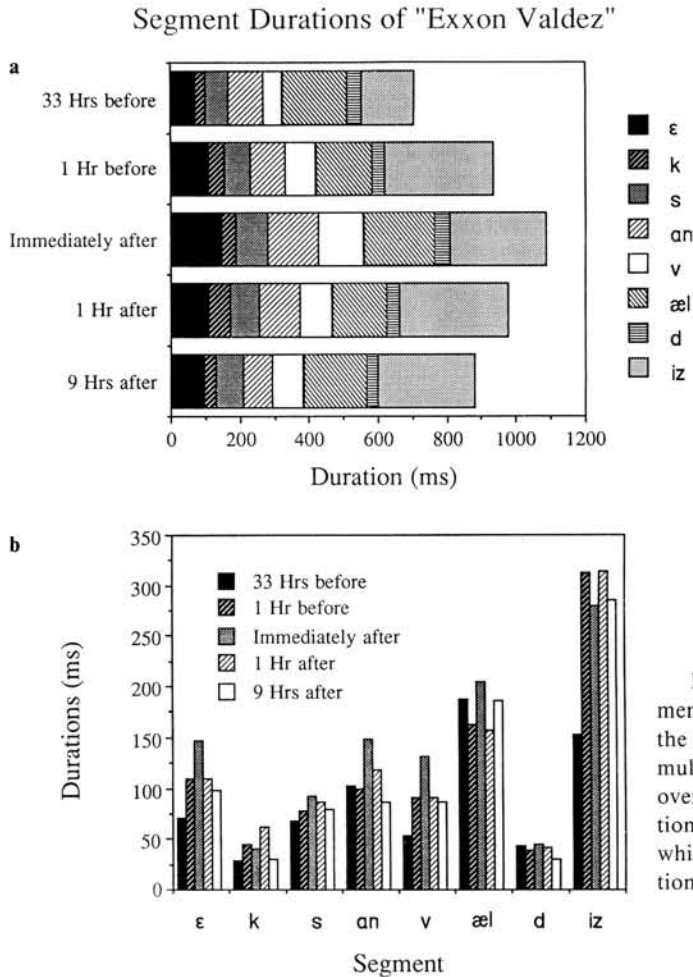


Fig. 4. Durations of speech segments in the phrase *Exxon Valdez* at the different times of recording. **a** Cumulative durations indicating the overall increase in duration. **b** Durations grouped by segments showing which segments had increased duration around the time of the accident.

samples recorded around the time of the accident than in the other samples on the NTSB tape.

One occurrence of the word *Valdez* occurred in the televised interview. This word was spoken in a discourse position which was comparable to that of *Exxon Valdez* in the NTSB recordings (utterance-initial position in a short sentence). Figure 5a compares the duration of *Valdez* in the inter-

view with the occurrences of this word in the NTSB recordings. This comparison suggests that the captain was speaking at his normal rate in the recording made 33 h before the accident, but more slowly in the recordings made around the time of the accident.

We also measured the duration of the phrase *thirteen and sixteen* which occurred in discourse-final position in three of the

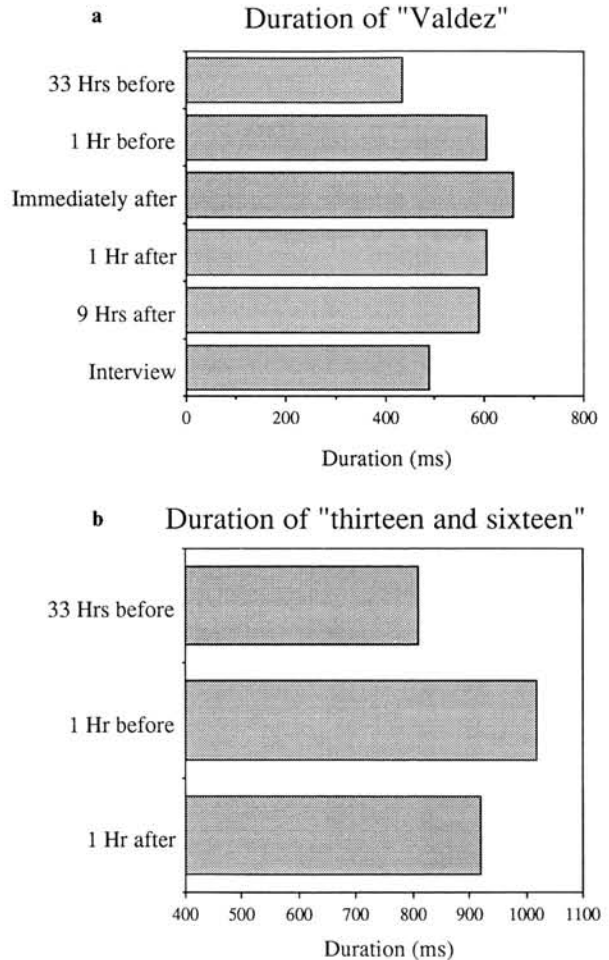


Fig. 5. **a** Duration of the word *Valdez* from the NTSB tapes (data is the same as that in fig. 4) compared with the same word produced in a similar discourse position in the televised interview. **b** Duration of the phrase *thirteen and sixteen* from recordings made at three times around the time of the accident.

recordings (33 h before the accident, 1 h before the accident, and 1 h after the accident). These measurements are shown in figure 5b. As with the durations of the phrase *Exxon Valdez*, these analyses indicate that Captain Hazelwood was speaking more slowly in the recordings made around the time of the accident than in the recording made 33 h before the accident.

Durational changes are perhaps the most reliable effects we have found in the NTSB

recordings, and they suggest that Captain Hazelwood was speaking more slowly than normal around the time of the accident. These changes in duration are consistent with the laboratory findings reported for speech produced while intoxicated, in noisy environments, under stress (for some subjects), and in simulations of the emotions fear, anger, and sorrow (table 2).

Figure 6a shows voice F_0 averaged across the phrase *Exxon Valdez* in each of the

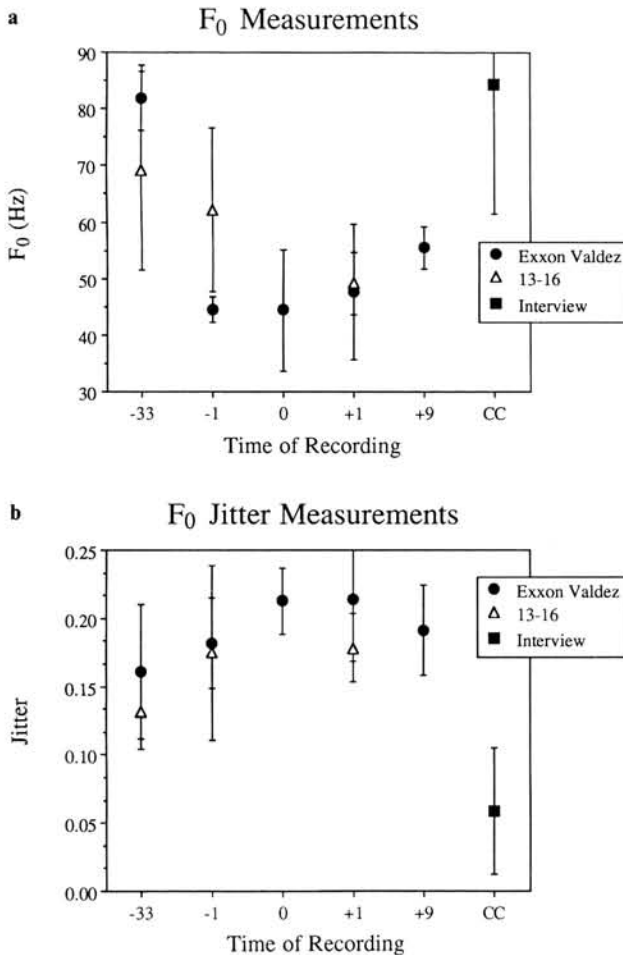


Fig. 6. **a** Average F_0 in *Exxon Valdez*, *thirteen and sixteen* and from one sentence in the televised interview as a function of time of recording. **b** F_0 jitter measurements from the same speech samples.

speech samples, the phrase *thirteen and sixteen* from three of the NTSB recordings, and one sentence from the televised interview. We took F_0 measurements from each of the four vowels in *Exxon Valdez* (which occurred at least twice in each of the NTSB recordings). We were not able to measure F_0 in all of the vowels in *thirteen and sixteen* because this phrase occurred in utterance-final position in the recordings and was

produced, in some cases, with quite low amplitude. Each point in figure 6 for *thirteen and sixteen* is based on measurements from at least two vowels. The last point in each panel shows data averaged across a sentence in the televised interview.

(The sentence was, 'I would say the same for the state of Alaska, they came after me, hammer and tong.')

The normal pitch detection algorithms

were unable to operate on the NTSB speech samples because of the amount of background noise; therefore, we modified an existing vocal jitter algorithm [see Pinto and Titze, 1990, for a recent review]. We adapted the existing technique by rectifying and low-pass filtering the signal (to remove high-frequency noise) before attempting to find successive pitch periods. The results of the algorithm were visually confirmed and then F_0 and jitter measures calculated. We calculated Davis's [1976] pitch perturbation quotient, which is the ratio of the 'average perturbation measured from the pitch period' and the average pitch period (pp. 51, 123).

As figure 6a shows, voice F_0 was dramatically lower in the samples recorded around the time of the accident. (F_0 as low as that seen here normally occurs only in a mode of vocal fold vibration called creak, or pulse register phonation. In English, this mode of vocal fold vibration is usually seen only at the ends of declarative sentences, although this varies somewhat from speaker to speaker.) This finding is consistent with previous studies of the effects of intoxication (for most subjects), stress (some subjects), and (portrayals of) sorrow. Also, this panel shows the average F_0 range in each speech sample. The different samples cannot be distinguished by their F_0 range (except perhaps the items from the recording made 9 h after the accident), but there was a trend for items near the time of the accident to have more F_0 jitter (fig. 6b). This finding is consistent with the studies listed in table 2, which found increases in jitter or standard deviation of F_0 , given the assumption above that increased jitter may produce an increase in SD F_0 (noise, workload, fear, anger, and intoxication).

In summary, the overall pattern of acoustic-phonetic changes which we have observed here is consistent (to varying degrees) with the findings of previous controlled laboratory studies of three effects on speech production: intoxication, stress, and sorrow. We will review here the ways in which these data are consistent with the effects of intoxication and will return to stress and sorrow in the conclusion. In listening to the recordings, we observed a number of gross misarticulations and segmental misarticulations around the time of the accident. We also found acoustic evidence in two of the recordings made near the time of the accident (0, +1) for a misarticulation of /s/. Finally, we found that Captain Hazelwood was speaking more slowly, and used a lower F_0 with more F_0 jitter around the time of the accident as compared with his speech 33 h before the accident and his speech in the televised interview.

Conclusions

We have presented *a priori* arguments which suggest that the effects of alcohol on speech production are, in principle, unique due to the local and global physiological effects of alcohol and the complexity of speech motor control. In a review of previous research on environmental and emotional effects of alcohol we found that this expectation was borne out within limits. Before discussing the possible explanations of the data reported here we will discuss three limitations on our ability to determine upon the basis of voice recordings whether Captain Hazelwood was intoxicated at the time of the accident.

First, there are gaps in the previous re-

search literature; both in research concerning the effects of alcohol on speech production and in research on other effects on speech production. For instance, we have reported here measurements of vocal jitter. This is the first time that vocal jitter measurements have been made in any study of the effects of alcohol on speech. We also noted several gaps in previous research on environmental and emotional effects on speech. For instance, we are not aware of any research which has attempted to explore the effects of fatigue on speech, or any research which explores the ways in which various environmental and/or emotional factors may interact in speech production. In the absence of these types of additional data, we cannot rule out a number of possible causes for the changes we have observed in Captain Hazelwood's speech.

Second, in addition to a lack of breadth in the existing knowledge, there is a lack of depth. There are no normative data on the effects of alcohol on speech production in the published literature. We do not know how general the effects summarized in table 1 are. Normative data are also unavailable for the effects summarized in table 2. This lack of data makes it difficult for us to offer reliable probabilistic statements such as, 'Captain Hazelwood had this pattern of changes and 95% of the people who exhibited this pattern were intoxicated while only 10% of fatigued speakers show this pattern.' Currently, statements of this type are based on studies which employed relatively small numbers of talkers.

Third, the recordings which we analyzed in the present case limited the type and quality of the acoustic measurements we could make. For instance, it would have been very informative to know whether the

captain was speaking more loudly or softly in the recordings near the time of the accident. This measure was not possible with the NTSB recordings because automatic gain control was used in the transmission equipment and the placement of the microphone in relation to the speaker's lips was (presumably) variable. Furthermore, the variability of the background noise made the calculation of long-term average spectra invalid, though Klingholz et al. [1988] found reliable changes in long-term average spectra when speakers were intoxicated. Our analysis of fricative spectra was also hampered by the presence of background noise and the frequency response characteristics of the transmission equipment. Finally, the complicated history of the recording made 33 h before the accident casts some doubt on the measurements taken from that recording. We have outlined the magnitude of error which may have resulted from this situation and have taken measurements from a televised interview to serve as another 'control' condition. Still, this extra link in the history of the recording introduces an additional source of error that would not have existed if the original Coast Guard recording had not been erased.

With these limitations in mind, we now turn to several possible explanations of the pattern of changes found in Captain Hazelwood's speech. Considering only the suprasegmental effects listed in table 2, there are three effects which are consistent with the pattern of changes observed in Captain Hazelwood's speech. Only stress, sorrow, and intoxication have been found to involve both increased duration and decreased F_0 .

One could easily suppose that the accident was a stress-inducing factor and thus

possibly that the pattern of changes observed here are the result of psychological stress. There are three problems with this hypothesis. First, there are no reports in the literature of an individual displaying both F_0 decrease and duration increase in a stressful situation. For the speakers studied by Streeter et al. [1983] F_0 and duration were positively correlated, for one speaker F_0 and duration increased while for the other F_0 and duration decreased. Captain Hazelwood showed a negative correlation, F_0 decreased and duration increased. Other studies of stress have also found a negative correlation, but in the opposite direction (F_0 increase and duration decrease). Second, in all studies of stress which have included measurements of F_0 jitter, jitter decreased in the stress condition. This may not be a very important consideration because these studies also found an increase in F_0 , and there appears to be a correlation between measured jitter and F_0 . Third, the pattern of changes which we observed in the captain's speech were present in the recordings made 1 h before the accident. Therefore, if we assume that the observed changes were due to increased stress, we also must assume that this stress was not a result of the accident because the phonetic changes appeared before the accident.

We could also suppose the captain responded to the accident with sorrow or depression, but again this does not account for the pattern of changes observed here because the changes in duration and F_0 appeared before the accident occurred. Therefore, if we wish to explain the pattern of results in terms of increased stress or sorrow we must assume that the event(s) which precipitated these conditions occurred prior to the accident.

There are two other effects that have not been previously studied which nonetheless are possible explanations of the phenomena observed in the captain's speech. We have no acoustic data on the changes in speech which could be expected when the speaker has recently been awakened. As mentioned earlier, our subjective impression is that there may be some similarities between speech produced in this situation and speech produced while intoxicated. One problem with this explanation of the pattern of changes found in the captain's speech is that the pattern persisted over a period of a least 10 h. It seems unlikely that the effects of being recently awakened would extend over such a long time, especially in these circumstances.

(Note that this argument about the persistence of the effects does not apply to the intoxication explanation. At an absorption rate of 0.015%/h [Wallgren and Barry, 1970] it takes over 10 h for blood alcohol level to go from 0.2% [well below the level for a fatal dose, which is about 0.4%] to 0.05%.)

We pointed out earlier that there are also some subjective similarities between the effects of fatigue and intoxication on speech production (we are not talking here about vocal fatigue, but rather general fatigue). So, it might be possible to consider the pattern of changes found here to be the results of fatigue. There is one aspect of the data presented above on Captain Hazelwood's speech which does not fit this explanation of the changes. The duration and F_0 measurements from the recording made 9 h after the accident (when we would expect the greatest amount of fatigue) are more similar to the values seen in the 'control' recordings (-33, CC) than are the measurements made

from the recordings made within 1 h of the accident. If we assume that the changes observed here were caused by fatigue, we must also assume that the effects of this fatigue diminished over time.

The only inconsistency in considering intoxication as an explanation of the changes observed in Captain Hazelwood's speech has to do with the data on his pronunciation of /s/. Why was there no evidence of an /s/-to-/ʃ/ change in the recording made 1 h before the accident?

There is evidently no simple explanation of the pattern of changes observed in these recordings. Considering the data as a whole, intoxication seems to be the simplest explanation, but these data are, in the final analysis, inconclusive. The evidence is necessarily inconclusive because one can imagine a number of complicated scenarios involving combinations of factors (some of which have unknown consequences for speech production) which might account for the observed changes. For this reason, it is not currently possible to arrive at a definitive conclusion concerning the physical state of the speaker upon the basis of voice recordings. Laboratory studies specifically attempt to eliminate complex interactions between factors by holding all conditions constant except the independent variable, and while results from controlled studies have some utility in evaluating speech patterns in real-life situations, it is important to keep in mind the limitations which necessarily exist in making inferences from controlled studies to situations in real life which involve a wide variety of unknown conditions. However, with additional research it may be possible to determine the limits and utility of this application of acoustic-phonetic analysis.

References

- Anttila, R.: An introduction to historical and comparative linguistics (MacMillan, New York 1972).
- Berry, M.S.; Pentreath, V.W.: The neurophysiology of alcohol; in Sandler, *Psychopharmacology of alcohol*, pp. 43–72 (Raven Press, New York 1980).
- Brenner, M.; Shipp, T.: Voice stress analysis; in *Mental-state estimation 1987*, pp. 363–376. NASA Conference Publication 2504 (1988).
- Brenner, M.; Shipp, T.; Doherty, E.T.; Morrissey, P.: Voice measures of psychological stress: laboratory and field data; in Titze, Scherer, *Vocal fold physiology, biomechanics, acoustics, and phonatory control*, pp. 239–248 (Denver Center for the Performing Art, Denver 1985).
- Borden, G.J.; Harris, K.S.: *Speech science primer: physiology, acoustics and perception of speech*; 2nd ed. (Williams & Wilkins, Baltimore 1984).
- Davis, S.B.: Computer evaluation of laryngeal pathology based on inverse filtering of speech. *SCRL Monogr.* 13 (1976).
- Dunker, E.; Schlosshauer, B.: Irregularities of the laryngeal vibratory pattern in healthy and hoarse persons; in Brewer, *Research potentials in voice physiology*, pp. 151–184. International Conference at Syracuse, 1961.
- Griffin, G.R.; Williams, C.E.: The effects of different levels of task complexity on three vocal measures. *Aviation Space env. Med.* 58: 1165–1170 (1987).
- Hansen, J.H.L.: Analysis and compensation of stressed and noisy speech with application to robust automatic recognition; PhD diss. Georgia Institute of Technology (1988).
- Hellekant, G.: The effect of ethyl alcohol on non-gustatory receptors of the tongue of the cat. *Acta physiol. scand.* 65: 243–250 (1965).
- Jakobson, R.: *Kindersprache, Aphasie und allgemeine Lautgesetze* (Uppsala 1941).
- Klatt, D.H.: Linguistic uses of segmental duration in American English: acoustic and perceptual evidence. *J. acoust. Soc. Am.* 59: 1208–1221 (1976).
- Klingholz, F.; Penning, R.; Liebhardt, E.: Recognition of low-level alcohol intoxication from speech signal. *J. acoust. Soc. Am.* 84: 929–935 (1988).

- Lehiste, I.: *Suprasegmentals* (MIT Press, Cambridge 1970).
- Lester, L.; Skousen, R.: The phonology of drunkenness; in Bruck, Fox, LaGaly, *Papers from the parasession on natural phonology*, pp. 233–239 (Chicago Linguistic Society, Chicago 1974).
- Lindblom, B.: Economy of speech gestures; in MacNeilage, *The production of speech*, pp. 217–245 (Springer, New York 1983).
- Lisker, L.; Abramson, A. D.: A cross-language study of voicing in initial stops: acoustic measurements. *Word* 20: 384–422 (1964).
- Moore, T.J.; Bond, Z.S.: Acoustic-phonetic changes in speech due to environmental stressors: implications for speech recognition in the cockpit. 4th Annu. Symp. on Aviation psychol., 1987, pp. 26–30.
- Pinto, N.B.; Titze, I.R.: Unification of perturbation measures in speech signals. *J. acoust. Soc. Am.* 87: 1278–1289 (1990).
- Pisoni, D.B.; Bernacki, R.H.; Nusbaum, H.C.; Yuchtman, M.: Some acoustic-phonetic correlates of speech produced in noise. *Proc. IEEE ICASSP*, pp. 1581–1584 (1985).
- Pisoni, D.B.; Hathaway, S.N.; Yuchtman, M.: Effects of alcohol on the acoustic-phonetic properties of speech. *Alcohol. accidents and injuries. Special Paper P-173*, pp. 131–150 (Society of Automotive Engineers, Pittsburgh 1986).
- Pisoni, D.B.; Martin, C.S.: Effects of alcohol on the acoustic-phonetic properties of speech: perceptual and acoustic analyses. *Alcoholism clin. exp. Res.* 13: 577–587 (1989).
- Scherer, K.R.: Vocal indicators of stress; in Darby, *Speech evaluation in psychiatry*, pp. 171–188 (Grune & Stratton, New York 1981).
- Sobell, L.; Sobell, M.: Effects of alcohol on the speech of alcoholics. *J. Speech Hear. Res.* 15: 861–868 (1972).
- Sobell, L.; Sobell, M.; Coleman, R.: Alcohol-induced dysfluency in nonalcoholics. *Folia phoniat.* 34: 316–323 (1982).
- Stevens, K.N.: The quantal nature of speech: evidence from articulatory-acoustic data; in David, Denes, *Human communication: a unified view* (McGraw-Hill, New York 1972).
- Streeter, L.A.; MacDonald, N.H.; Apple, W.; Krauss, R.M.; Galotti, K.M.: Acoustic and perceptual indicators of emotional stress. *J. acoust. Soc. Am.* 73: 1354–1360 (1983).
- Subtelny, J.D.; Oya, N.; Subtelny, J.D.: Cineradiographic study of sibilants. *Folia phoniat.* 24: 30–50 (1972).
- Summers, W.V.; Pisoni, D.B.; Bernacki, R.H.: Effects of cognitive workload on speech production: acoustic analyses. *Res. on Speech Perception Prog. Rep. No. 15*, pp. 485–502 (Indiana University Department of Psychology, Bloomington 1989).
- Summers, W.V.; Pisoni, D.B.; Bernacki, R.H.; Pedlow, R.I.; Stokes, M.A.: Effects of noise on speech production: acoustic and perceptual analyses. *J. acoust. Soc. Am.* 84: 917–928 (1988).
- Trojan, F.; Kryspin-Exner, K.: The decay of articulation under the influence of alcohol and paraldehyde. *Folia phoniat.* 20: 217–239 (1968).
- Villiers, J.G. de; Villiers, P.A. de: *Language acquisition* (Harvard University Press, Cambridge 1978).
- Wallgren, H.; Barry, H.: *Actions of alcohol*, vol. 1 (Elsevier, Amsterdam 1970).
- Williams, C.E.; Stevens, K.N.: Emotions and speech: some acoustical correlates. *J. acoust. Soc. Am.* 52: 1238–1250 (1972).
- Williams, C.E.; Stevens, K.N.: Vocal correlates of emotional states; in Darby, *Speech evaluation in psychiatry*, pp. 221–240 (Grune & Stratton, New York 1981).

Received: June 25, 1990

Accepted: September 13, 1990

Keith Johnson
Department of Linguistics
UCLA
Los Angeles, CA 90024–1543 (USA)