

Maurer School of Law: Indiana University Digital Repository @ Maurer Law

Articles by Maurer Faculty

Faculty Scholarship

1991

Novel Scientific Evidence of Intoxication: Acoustic Analysis of Voice Recordings from the Exxon Valdez

J. Alexander Tanford

Indiana University Maurer School of Law, tanford@indiana.edu

David B. Pisoni

Indiana University - Bloomington

Keith Johnson

University of California - Los Angeles

Follow this and additional works at: <http://www.repository.law.indiana.edu/facpub>

 Part of the [Criminal Law Commons](#), and the [Evidence Commons](#)

Recommended Citation

Tanford, J. Alexander; Pisoni, David B.; and Johnson, Keith, "Novel Scientific Evidence of Intoxication: Acoustic Analysis of Voice Recordings from the Exxon Valdez" (1991). *Articles by Maurer Faculty*. Paper 677.

<http://www.repository.law.indiana.edu/facpub/677>

This Article is brought to you for free and open access by the Faculty Scholarship at Digital Repository @ Maurer Law. It has been accepted for inclusion in Articles by Maurer Faculty by an authorized administrator of Digital Repository @ Maurer Law. For more information, please contact wattn@indiana.edu.

NOVEL SCIENTIFIC EVIDENCE OF INTOXICATION: ACOUSTIC ANALYSIS OF VOICE RECORDINGS FROM THE EXXON VALDEZ

J. Alexander Tanford*, David B. Pisoni & Keith Johnson*****

Author's Notes: Part of this article reports original research conducted under the direction of the second and third authors. The initial research was supported by a contract to Indiana University from General Motors Research Laboratories. The specific analyses of voice recordings of Captain Joseph Hazelwood were conducted by them at the request of the National Transportation Safety Board, and are based on tapes and data supplied by the NTSB. The second author may be called as a witness in some of the lawsuits pending against the Exxon Corporation. The opinions expressed in this article concerning whether this evidence meets the legal standards of reliability and admissibility are those of the first author, who is not affiliated with the Speech Research Laboratory and has not participated in either the initial research nor the analysis of the Exxon Valdez tapes.

I. INTRODUCTION

On March 24, 1989, the oil tanker Exxon Valdez ran aground in Prince William Sound, causing the worst accidental oil spill in history. Captain Joseph J. Hazelwood was in command of the ship, although he was not on the bridge at the time of the accident. Almost immediately, rumors and allegations surfaced suggesting that Hazelwood was intoxicated, but no hard evidence came to light. A captain at sea is a long way from the nearest police officer with a breathalyzer. Hazelwood denied having been intoxicated, and an Alaska jury acquitted him of the charge. There seems to be no way to prove whether or not Hazelwood really had been drinking. Or is there?

* Professor of Law and Ira C. Batman Fellow, Indiana University — Bloomington.

** Professor of Psychology and Cognitive Science & Director, Speech Research Laboratory, Indiana University — Bloomington.

*** Phonetics Laboratory, Department of Linguistics, University of California at Los Angeles.

Recent work in the Speech Research Laboratory at Indiana University's Psychology Department suggests that intoxication can be detected by acoustic-phonetic analyses of a suspect's voice.¹ The National Transportation Safety Board (NTSB), the federal agency charged with investigating the Exxon Valdez accident, made available to the authors copies of tape recordings of conversations which took place between Captain Hazelwood and the Coast Guard from before, during, and after the accident.² Analyses of these tapes suggest that Hazelwood was indeed intoxicated when his ship ran aground.³ This novel application of speech science techniques for measuring the effects of alcohol on speech has never, to the knowledge of the authors, been attempted in a civil or criminal case.

Our analyses raise several obvious questions: Are the results reliable? How certain can we be that Captain Hazelwood was intoxicated? And, if this new testing procedure, based on voice recordings, can indeed determine that someone was under the influence of alcohol, are the results admissible in court?

The answers to these questions have implications beyond their obvious application to determining fault in the more than three hundred pending lawsuits concerning the Alaska oil spill.⁴ In many cases, defendants may be far away from the nearest breathalyzer or may refuse to submit to blood-alcohol tests. In the absence of reliable, objective evidence of whether a defendant had been drinking, courts must rely on a presumption of guilt based on a defendant's refusal to take a test,⁵ admit the speculative opinions of witnesses, or let the jury guess after listening to tape recordings.⁶ If a new measurement procedure exists that can, at least in some cases, provide objective indications of a person's physical state or condition,

¹ David B. Pisoni & Christopher S. Martin, *Effects of Alcohol on the Acoustic-Phonetic Properties of Speech: Perceptual and Acoustic Analysis*, 13 ALCOHOLISM: CLINICAL & EXPERIMENTAL RES. 577 (1989); Keith Johnson et al., *Do Voice Recordings Reveal Whether a Person is Intoxicated? A Case Study*, 47 PHONETICA 215 (1990).

² Cf. *New York Times Co. v. Nat'l Aeronautics and Space Admin.*, 920 F.2d 1002 (D.C. Cir. 1990) (voice recordings of deceased Challenger crew held within a Freedom of Information Act exemption; their release could unfairly invade a right of privacy).

³ Captain Hazelwood denies that he was intoxicated. His attorney, Michael G. Chalos, has called the analytical method discussed here "voodoo stuff." Katherine Bishop, *Leaps of Science Create Quandaries on Evidence*, N.Y. TIMES, April 6, 1990, at B6.

⁴ As of March 13, 1991, over three hundred lawsuits against Exxon were pending that are not affected by Exxon's settlement with the Environmental Protection Agency. Keith Schneider, *Exxon to Pay \$100 Million Fine and Plead Guilty in Valdez Spill*, N.Y. TIMES, March 13, 1991, at A1.

⁵ See *South Dakota v. Neville*, 459 U.S. 553, 564 (1983) (approving the admissibility of a refusal to submit to a blood-alcohol test).

⁶ See *Pennsylvania v. Muniz*, 110 S. Ct. 2638, 2642 (1990) (jury heard video and audio tapes and testimony about field sobriety tests).

then courts could more reliably convict the guilty and acquit the innocent in cases where no chemical blood test results exist.

In this article, we describe this new testing procedure, using the analyses performed on Captain Hazelwood's voice as an example. We then discuss whether the results should be admissible under the rules governing novel scientific evidence. We conclude that the kinds of acoustic-phonetic analyses described in this article produce reliable and relevant evidence that should be admitted when supported by proper expert testimony. We do not claim that speech analyses conclusively prove that Captain Hazelwood was intoxicated when the Exxon Valdez ran aground. Rather, we believe that the analyses of Hazelwood's voice produce data consistent with intoxication. They therefore should be admitted into evidence and considered by the jury along with other relevant information.

II. THE SCIENTIFIC EVIDENCE: DETERMINING INTOXICATION FROM VOICE RECORDINGS

Alcohol is generally considered to be a central nervous system depressant. Significant blood concentrations of alcohol have been found to impair coordination, reflexes and nerve transmissions.⁷ This kind of loss of motor control would seem naturally to affect speech. Indeed, controlled laboratory studies have demonstrated that alcohol produces three kinds of changes in speech production: gross effects, segmental effects and suprasegmental effects. Examples of each are listed in Table 1.

Gross effects in speech production involve the alteration of entire words. Alcohol impairs a speaker's ability to retrieve words from memory and utter them in the proper sequence. A speaker who has consumed alcohol may revise, omit or interject words, sounds, or phrases.⁸ One common example of a gross effect is the reversal of two words in a sentence (a spoonerism), e.g., "*work* is the curse of the *drinking* class."⁹

Segmental effects, on the other hand, involve the misarticulation of specific speech sounds, notably the phonemes /r/, /l/, /s/ and /ts/. These mispronunciations are easily detected in spontane-

⁷ See M.S. Berry & V.W. Pentreath, *The Neurophysiology of Alcohol*, in *PSYCHOPHARMACOLOGY OF ALCOHOL* 43-72 (1980); Pisoni & Martin, *supra* note 1, at 577.

⁸ See Linda C. Sobell & Mark B. Sobell, *Effects of Alcohol on the Speech of Alcoholics*, 15 *J. SPEECH HEARING RES.* 861, 866 (1972); Linda C. Sobell et al., *Alcohol-induced Dysfluency in Nonalcoholics*, 34 *FOLIA PHONIATRICA* 316 (1982). Gross effects may be hard to recognize in spontaneous speech because the speaker's intended utterance is unknown.

⁹ GLORIA J. BORDEN & KATHERINE S. HARRIS, *SPEECH SCIENCE PRIMER: PHYSIOLOGY, ACOUSTICS, AND PERCEPTION OF SPEECH* 56 (2d ed. 1984).

ous speech if one can compare examples of a person's intoxicated speech to that person's speech while sober.¹⁰ Although some segmental effects may accompany any kind of loss of motor control,¹¹ the substitution of an /sh/ sound for /s/ seems to be unique to loss of control caused by alcohol.¹²

Finally, suprasegmental effects involve changes in the duration, pitch and amplitude of speech. Intoxicated speakers talk more slowly, often use a lower mean pitch, and display greater pitch variation than they do when they are sober.¹³ When these changes in speaking rate and pitch can be quantified, they are usually the most salient indication of alcohol impairment. Differences in rate and pitch of an intoxicated speaker can be measured objectively using conventional digital signal processing techniques.¹⁴ The resulting measurements may then be easily compared to similar measurements made from samples of the speaker's sober speech.¹⁵ Two other common suprasegmental effects are vowel lengthening¹⁶ and lengthening of consonants in unstressed syllables.¹⁷

In trying to determine if Captain Hazelwood had consumed alcohol, we applied speech analysis techniques to five samples of his speech culled from audio tapes provided by the NTSB. According to the NTSB, these recordings were made thirty-three hours before the accident, one hour before the accident, immediately after the accident, one hour later, and nine hours after the accident. All

¹⁰ See Leland Lester & Royal Skousen, *The Phonology of Drunkenness*, in PAPERS FROM THE PARASESSION ON NATURAL PHONOLOGY 233 (Anthony Bruck et al. eds., 1974); F. Trojan & K. Kryspin-Exner, *The Decay of Articulation Under the Influence of Alcohol and Paraldehyde*, 20 FOLIA PHONIATRICA 217 (1968); David B. Pisoni et al., *Effects of Alcohol on the Acoustic-Phonetic Properties of Speech*, in ALCOHOL, ACCIDENTS AND INJURIES 131 (Soc'y of Automotive Engineers, Inc. Technical Paper No. 860361, 1986).

¹¹ For example, common speech errors include consonant reversals ("a two-sen pet" instead of "a two-pen set") and vowel reversals ("fool the pill" instead of "fill the pool"). BORDEN & HARRIS, *supra* note 9, at 57.

¹² See Johnson et al., *supra* note 1, at 219.

¹³ Pisoni & Martin, *supra* note 1, at 582-83. Cf. Sobell et al., *supra* note 8, at 320 (finding no significant effects on fundamental frequency at moderately high levels of intoxication (0.10%)); Trojan & Kryspin-Exner, *supra* note 10, at 222 (effects on pitch varied).

¹⁴ Standard operating procedures were used in analyzing Captain Hazelwood's speech. The voice recordings were low-pass filtered at 9.6 kHz and digitized at a 20-kHz sampling rate through a 12-bit Analog-to-Digital (A/D) converter. A digital waveform editor was used with a PDP 11/34 minicomputer to edit all speech samples into separate digital files. See Pisoni & Martin, *supra* note 1, at 577-78.

¹⁵ See Pisoni & Martin, *supra* note 1; F. Klingholtz et al., *Recognition of Low-Level Alcohol Intoxication from Speech Signals*, 84 J. ACOUSTIC SOC'Y AM. 929 (1988); Sobell et al., *supra* note 8.

¹⁶ See Pisoni & Martin, *supra* note 1, at 581.

¹⁷ See Lester & Skousen, *supra* note 10, at 233-34.

Table 1
Effects of alcohol on speech.

Gross effects	Word/phrase/syllable interjection. Word omission. Word revision. Broken suffixes.
Segmental effects	Misarticulation of "r" and "l". "S" becomes "sh". Final devoicing, e.g., "iz" -> "is". Deaffrication, e.g., "church" -> "shursh".
Suprasegmental effects	Reduced speaking rate. Decreased amplitude. Mean change in pitch range. Increase in pitch variability.

taped communications contained the phrase "Exxon Valdez." This utterance was the focus of our quantitative analyses because it was the same phrase used several times, and it contained tokens of the fricative consonant /s/.

We first looked for gross effects. Gross effects usually are difficult to measure in spontaneous speech, because the listener does not know what word was intended by the speaker. Nevertheless, we found several examples of gross errors where the intended word was obvious, or the speaker himself corrected the mistake. These are summarized in Table 2.

We also found several segmental errors in Captain Hazelwood's speech. We paid particular attention to the /s/ sounds in the phrase "Exxon Valdez." Although Hazelwood pronounced Exxon correctly the day before the accident, he misarticulated it as "ekshon" around the time of the accident. This subjective perception was confirmed by spectral analysis.¹⁸ Additionally, the recordings made

¹⁸ Spectral analysis techniques are commonly used to measure the distribution of

Table 2
Summary of Gross Effects Found in NTSB Tape

First word used	Revision
1. Exxon Ba...	Exxon Valdez
2. Departed	Disembarked
3. I	We'll
4. Columbia Gla	Columbia Bay

near the time of the accident reveal several misarticulations of /r/ and /l/ in the words "northerly," "little," "drizzle," and "visibility". Finally, the authors found several examples of final devoicing, as /z/ in Valdez became an /s/ ("valdes"). These findings suggest that at the time his ship ran aground, Captain Hazelwood was having difficulty with the fine motor control used to produce these sounds. The findings do not, however, prove that this loss of motor control was caused by alcohol consumption.

The most salient indication that a loss of motor control is due to alcohol consumption is the prevalence of suprasegmental effects. We focused on speaking rate and voice fundamental frequency.¹⁹ We measured the duration of the speech segments in the phrase "Exxon Valdez" in each sample, taking the average of at least two occurrences of the phrase in each time period. The results, summarized in Table 3, show that it took Captain Hazelwood approximately 50% longer to say the phrase "Exxon Valdez" at the time of the accident than the day before.

Speech theory predicts that changes in duration will be most noticeable in vowel segments²⁰ and unstressed consonant segments.²¹ Looking at the /E/ and /on/ in Exxon and the initial /V/

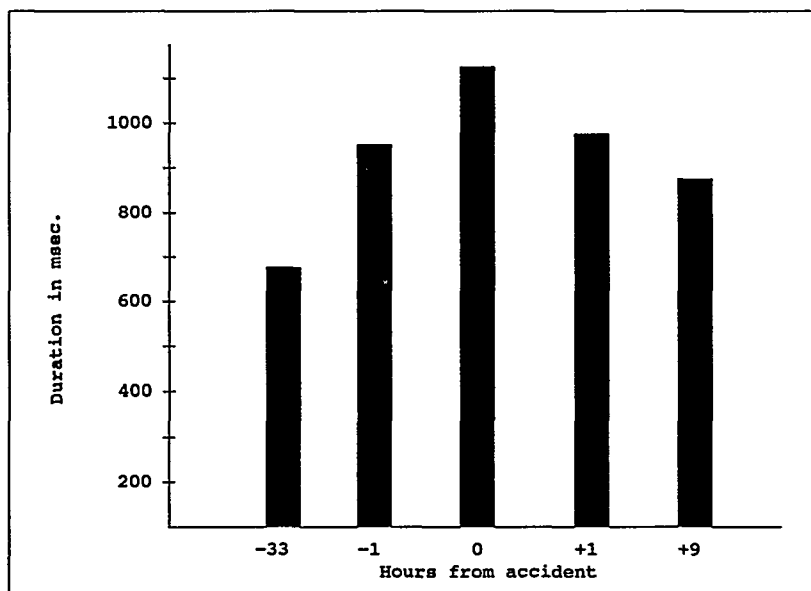
energy at different frequencies as a function of time. The results are often displayed as speech spectrograms, or "voiceprints," in order to reveal the dynamic time-varying nature of speech. See JAMES L. FLANAGAN, *SPEECH ANALYSIS SYNTHESIS AND PERCEPTION* 141-63 (2d ed. 1972).

¹⁹ Because the communication equipment had automatic gain controls and the distance between the speaker and the microphone probably varied, we could not measure changes in the amplitude of the speech.

²⁰ See Pisoni & Martin *supra* note 1, at 581.

²¹ See Lester & Skousen, *supra* note 10.

Table 3
Cumulative Duration of Speech Segments in "Exxon Valdez"



in Valdez, the slowed speech is particularly apparent. These results are summarized in Table 4.

Finally, we calculated voice fundamental frequency (pitch) across the phrase "Exxon Valdez." For each of the five relevant time periods, we measured the pitch of all four vowel sounds in two productions of "Exxon Valdez", and averaged the eight measurements. Voice pitch was dramatically lower in the samples recorded around the time of the accident. Variability of fundamental frequency was correspondingly greater in the voice samples taken at the time of the accident. Lowered pitch and increased variability are characteristic of alcohol-impaired speech, so this evidence also suggests that Captain Hazelwood had been consuming alcohol at the time his ship ran aground. These data are summarized in Tables 5 and 6.

The acoustic-phonetic differences found in the speech samples supplied by the NTSB are consistent with the findings of controlled laboratory studies on the effects of alcohol on speech.²² However, they do not *prove* that Hazelwood was drunk. In fact, unless other obvious explanations for the pattern of changes can be discounted,

²² See Klingholtz et al., *supra* note 15; Lester & Skousen, *supra* note 10; Pisoni et al., *supra* note 10; Pisoni & Martin, *supra* note 1; Sobell & Sobell, *supra* note 8; Sobell et al., *supra* note 8; Trojan & Kryspin-Exner, *supra* note 10.

the pattern is not even strong evidence of intoxication. Speech is affected by the physical state of the speaker, especially when that state is characterized by stress or fatigue. Measurements made from tape recordings can be affected by malfunctions or variations in the recording equipment. In the Hazelwood case, however, these alternative explanations can be discounted.

A. THE PHYSICAL STATE OF THE SPEAKER

Joseph Hazelwood would undoubtedly have been under some psychological stress following the Exxon Valdez accident, and would logically have become increasingly fatigued as the hours wore on. However, neither of these two factors appears to explain the specific changes in his speech found on the tapes.

Stress might be expected to affect Hazelwood's speech immediately after, one hour after and nine hours after the accident, but it could hardly have affected his speech one hour *before* the accident occurred. Yet, both segmental and suprasegmental effects are just as robust one hour before the accident as one hour afterwards. In addition, previous studies on speech production in stressful environments show that stress and alcohol affect speech in *opposite* ways. Stress causes pitch and speaking rate to increase rather than decrease.²³ Psychological states similar to stress, such as fear and anger, likewise produce increases in pitch and rate.²⁴ Alcohol, on the other hand, causes fundamental frequency (pitch) and speaking rates to decrease. In Hazelwood's case, both measures decreased.

Fatigue may cause changes in speech production which are similar to those caused by intoxication. Surprisingly, speech scientists have conducted little controlled research on the effects of fatigue on speech. In the absence of scientific data concerning the effects of fatigue on speech production, it is reasonable (and conservative) to assume that fatigue produces effects which are similar to the effects of intoxication: a general lowering of arousal, slower speaking rate, and lower voice pitch. It also seems reasonable to assume that a person's level of fatigue increases over time without sleep.

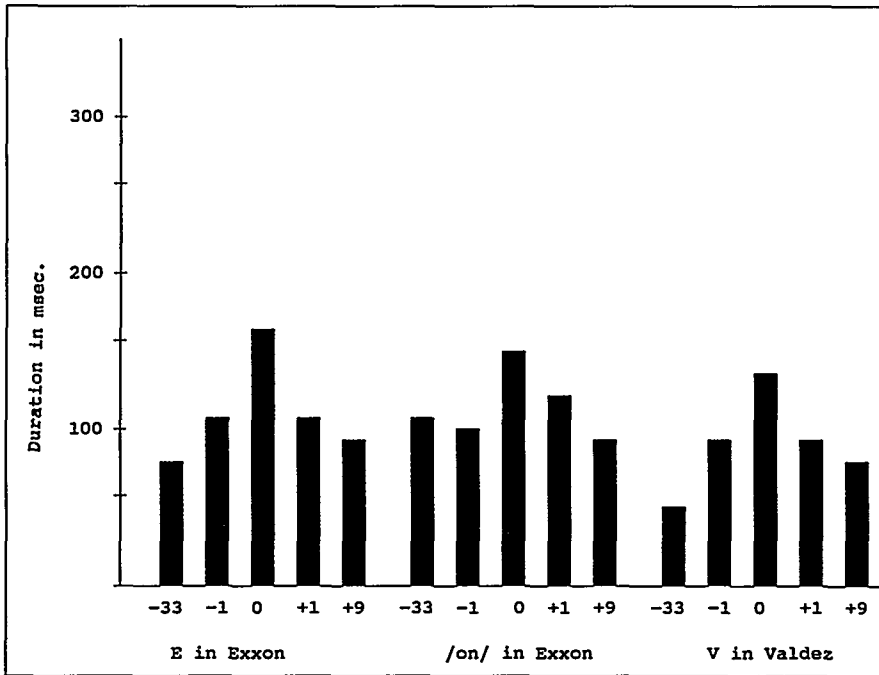
Given these assumptions, the pattern of phonetic changes seen in Captain Hazelwood's speech cannot be attributed to simple fa-

²³ See Malcolm Brenner & Thomas Shipp, *Voice Stress Analysis*, in MENTAL STATE ESTIMATION 363 (NASA Conf. Pub. #2504, 1988).

²⁴ See JOHN HANSEN, ANALYSIS AND COMPENSATION OF STRESSED AND NOISY SPEECH WITH APPLICATION TO ROBUST AUTOMATIC RECOGNITION, (University Microfilms 1988) (fear, anger); Carl E. Williams and Kenneth N. Stevens, *Emotions and Speech: Some Acoustical Correlates*, 52 J. ACOUSTICAL SOC'Y AM. 1238, 1248-49 (1972) (anger).

²⁵ See *supra*, tables 3-6.

Table 4
Duration of Vowel Sounds and Consonants in
Unstressed Syllables in "Exxon Valdez"



tigue. He spoke more slowly and with lower fundamental frequency at the time of the accident than nine hours later. Indeed, his speech nine hours after the accident was similar to his speech the day before. If the changes could be attributed to fatigue, one would expect that fatigue-induced effects would be greatest nine hours after the accident, by which time Captain Hazelwood had been deprived of sleep for twenty-two hours. This expectation was not borne out by the data.²⁵

B. THE MECHANICAL STATE OF THE RECORDINGS

Measurements of pitch and duration of sounds are sensitive to fluctuations in the speed of tape recordings. If an audio tape is played at a slower speed, the voice will sound lower in pitch, and speech sounds will seem to have taken longer to articulate. These effects are similar to those caused by alcohol impairment.

To some extent, the risk of erratic tape speeds is reduced as long as both the control recording (in which the speaker is presumed sober) and the test recording (in which the state of the speaker is in question) were recorded and played on the same

Table 5
Average Pitch of Voice in "Exxon Valdez"

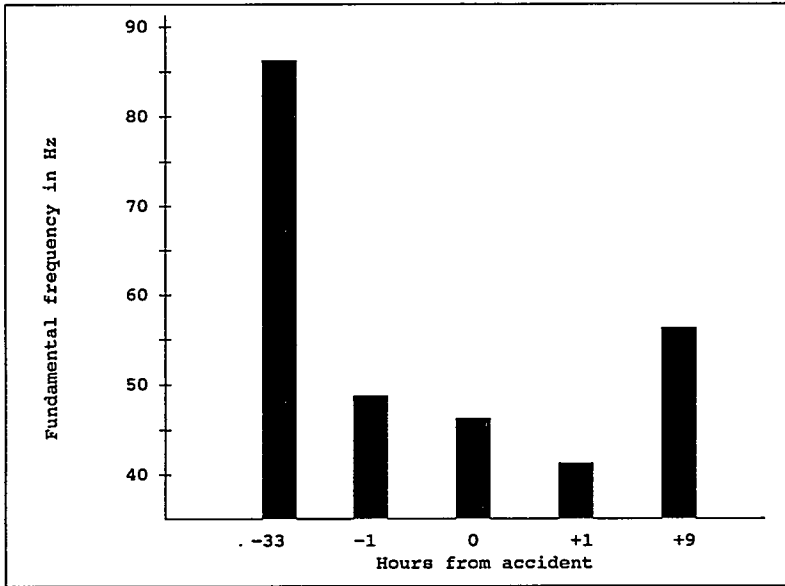
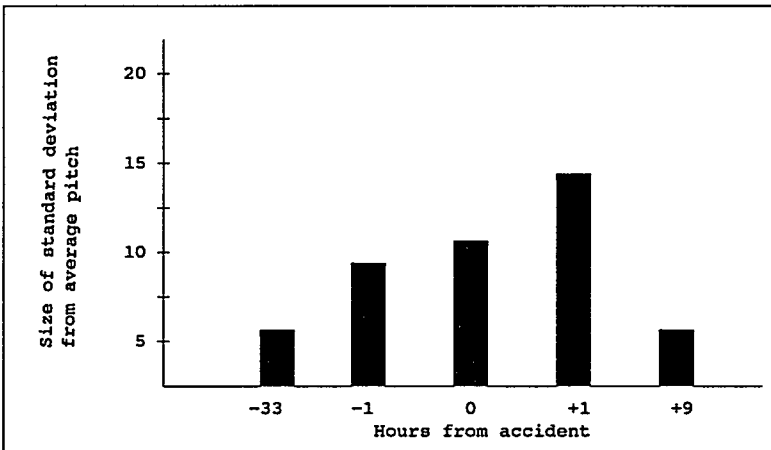


Table 6
Variation in Pitch of Voice in "Exxon Valdez"



equipment. The test is a comparative one that does not depend on the *true* pitch and duration being known. A more reliable control is to measure some sound other than the suspect's voice that appears on both tapes. In the Exxon Valdez case, for example, Captain Hazelwood was talking to the same Coast Guard radio operator on the various tapes. Hazelwood's voice pitch and duration changed while the radio operator's did not. It is also possible to measure the average pitch of background noise on two tapes. This was done for the Valdez tapes. Again, the background noise did not significantly change, while Hazelwood's voice did. These observations reduce the likelihood that the effects measured were the result of mechanical problems in the audio tape recording.

C. OTHER FACTORS

Research on environmental and emotional effects on speech production has demonstrated that other factors also can cause suprasegmental effects similar to those caused by alcohol. A noisy environment has been shown to increase the variability of the fundamental frequency and to produce slower speech, but speech produced in a noisy environment tends to have a *higher* fundamental frequency rather than lower.²⁶ In another experiment, acceleration and vibration affected fundamental frequency and duration, but in a direction opposite to that caused by alcohol.²⁷ Several studies on mental workload with high cognitive demands (e.g., airline pilots, air traffic controllers) indicate that speech affected by the performance of a cognitively demanding task will display suprasegmental effects opposite to those observed with alcohol — higher average frequency and shorter duration.²⁸ Sorrow produces lower average frequency and slower speech, but seems to cause less rather than more pitch variability.²⁹

In short, there are no known environmental situations or emotional states that produce quite the same pattern of suprasegmental effects as alcohol impairment. These observations mean that, *in theory*, it is possible to classify changes observed across two samples of

²⁶ See HANSEN, *supra* note 24; Walter Van Summers et al., *Effects of Noise on Speech Production: Acoustic and Perceptual Analyses*, 84 J. ACOUSTIC SOC. AM. 917 (1988).

²⁷ See Thomas J. Moore & Zinny S. Bond, *Acoustic-Phonetic Changes in Speech Due to Environmental Stressors: Implications for Speech Recognition in the Cockpit*, 4 ANN. SYMPOSIUM AVIATION PSYCH. 26 (1987). The Moore and Bond study is preliminary at best; only two subjects were tested.

²⁸ The effects on frequency variability are mixed. See G.R. Griffin & Carl E. Williams, *The Effects of Different Levels of Task Complexity on Three Vocal Measures*, 58 AVIATION SPACE ENVIRONMENTAL MEDICINE 1165 (1987); Hansen, *supra* note 24.

²⁹ See HANSEN, *supra* note 24; Williams & Stevens, *supra* note 24, at 1249.

speech as more like the pattern found with alcohol-affected speech than with any other probable cause of impairment. There are, however, three caveats. First, it is not possible to give any kind of confidence rating to such a classification. There is not enough published laboratory data on individual differences which would allow the calculation of hit rates and false alarm rates for classifications based on these acoustic measures.³⁰ Second, there are some possible physiological effects on speech production which have not yet been adequately studied, such as fatigue, illness and being suddenly awakened. Any of these might conceivably produce effects similar to those measured after alcohol consumption. Third, no data have been gathered in more complex environments involving combinations of these stimuli.

D. SCIENTIFIC CONCLUSION

The ultimate conclusion to be drawn from our data is this: Analyses of audio tapes of Captain Hazelwood's speech cannot prove that he was alcohol-impaired at the time of the Exxon Valdez accident. Our analyses, however, provide objective tests for determining which of several possible explanations of Captain Hazelwood's physical state is the most likely. For example, if phonetic analyses revealed faster speech and higher fundamental frequency, we could attribute those changes to increased arousal caused by stress or anger, rather than intoxication. As it turns out, no such alternative explanation fits the observed pattern of changes in a simple way.

Analyses show a pattern of changes consistent with alcohol impairment. Captain Hazelwood's speech around the time of the accident, compared to his speech thirty-three hours earlier or nine hours later, is characterized by misarticulations of /r/ and /l/, changes from /s/ to /sh/, final devoicing of /iz/ to /is/, reduced speaking rate, lower fundamental frequency, increased pitch variability, and a number of word and syllable revisions. This pattern of segmental and suprasegmental effects is consistent with the pattern of alcohol-impaired speech measured in a controlled laboratory environment. It is not consistent with patterns of speech affected by fear, anger, noise, acceleration, vibration, or increased mental workload. The pattern is partially consistent with speech affected by stress or sorrow, but these changes were observed on the audio tape

³⁰ For example, it is not possible to offer reliable probabilistic statements such as: "Hazelwood had this pattern, and 95% of people who exhibit this pattern are intoxicated and only 10% of fatigued speakers show this pattern."

made one hour *before* the accident, when Hazelwood would probably not yet have been experiencing either stress or sorrow. The effects probably were not caused by mechanical problems affecting tape speed. Our analyses cannot rule out the possibility that Hazelwood's speech was affected in whole or part by fatigue, although logic and general theories of speech motor control suggest this is a less likely explanation than alcohol consumption. The question posed by our analyses, then, is whether this scientific evidence is admissible in court.

III. THE LEGAL STANDARD FOR ADMITTING NOVEL SCIENTIFIC EVIDENCE

The problem of separating reliable scientific evidence from quackery is not new. For centuries, courts have wrestled with the question of whether to admit the testimony of expert witnesses. Although courts have not always come to the right results,³¹ they have generally agreed on the procedural rule by which science was to be measured: the "general acceptance" test announced in *Frye v. United States*.³² Within the last ten years, however, the effectiveness of the *Frye* test as a screening device for scientific evidence has come under increasing attack, led by Professor Paul Gianelli's watershed article.³³

A. THE *FRYE* TEST

The *Frye* test prohibits the courtroom use of scientific evidence until it has gained "general acceptance" in relevant scientific fields. The D.C. Court of Appeals announced the test in a case which ruled inadmissible the test results from a systolic-pressure measurement device designed to detect lies. The court stated, without precedent or citation:

Just when a scientific principle or discovery crosses the line between

³¹ In *A Trial of Witches*, 6 HOWELL'S STATE TRIALS 687, 697 (1665), two women of Leystoff, County of Suffolk, were accused of bewitching two children. The evidence was conflicting, so the court sought the opinion of "Dr. Brown of Norwich, a person of great knowledge" concerning witchcraft. Dr. Brown was clearly of the opinion that the victims had been bewitched, and explained the procedure whereby the devil excited victims' humours through pins inserted by witches. The fact that the form of bewitchment appeared to be natural swooning fits he attributed to the villainous "subtilty [sic] of the devil." His testimony was believed, and the witches were convicted and executed.

³² 293 F. 1013, 1014 (D.C.Cir. 1923).

³³ Paul C. Gianelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half-Century Later*, 80 COLUM. L. REV. 1197 (1980). That his article has led the trend away from *Frye* is somewhat ironic, as Gianelli intended to criticize most of the deviations from the *Frye* test and urge that we preserve the spirit of *Frye* by developing a more workable version of that test. See *id.* at 1250.

the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.³⁴

This vague language probably was intended only to justify the exclusion of "lie detector" test results. Its negative tone expresses the presumption that scientific evidence will be *excluded* rather than admitted. This sets the *Frye* test at odds with the most fundamental principle of evidence — that all facts having rational probative value are presumed admissible.³⁵ Nevertheless, this language was widely adopted by courts as a general principle of exclusion, and is still the controlling law in some jurisdictions.³⁶ Courts have justified the adoption of a restrictive rule in part because of the fear that jurors are "easily overawed by conclusions voiced in court by articulate experts with impressive credentials."³⁷

The major weakness of the *Frye* test is that it fails to clearly distinguish between novel scientific evidence and reliable scientific evidence. The general acceptance test arises from a fear of science,³⁸ and therefore tends to exclude all new scientific evidence, whether or not it is reliable. Courts and commentators, in recognition of this problem, have increasingly criticized the *Frye* test on three grounds: it is bad science, it is bad law, and it is premised on bad psychology.³⁹

1. *Frye is Bad Science*

The *Frye* test was supposed to help judges distinguish scientifically reliable evidence from nonsense. The assumption that general acceptance of a procedure is synonymous with scientific accuracy, however, is dubious. This is especially true if courts rely on other judicial opinions to determine general acceptance. Testing proce-

³⁴ *Frye*, 293 F.2d at 1014.

³⁵ See 1 JOHN H. WIGMORE, EVIDENCE IN TRIALS AT COMMON LAW § 10 (Peter Tillers rev. ed. 1983).

³⁶ E.g., *United States v. Shorter*, 809 F.2d 54, 60 (D.C. Cir. 1987). See CHARLES T. McCORMICK, McCORMICK ON EVIDENCE 606 (Edward W. Cleary ed., 3d ed. 1984); Joseph G. Petrosinelli, Comment, *The Admissibility of DNA Typing: A New Methodology*, 79 GEO. L.J. 313, 323 (1990); Steven M. Egesdal, Note, *The Frye Doctrine and Relevancy Approach Controversy: An Empirical Evaluation*, 74 GEO. L.J. 1769, 1769 (1986).

³⁷ ANDRE A. MOENSSENS ET AL., SCIENTIFIC EVIDENCE IN CRIMINAL CASES 6 (3d ed. 1986).

³⁸ See J. Alexander Tanford & Sarah Tanford, *Better Trials Through Science: A Defense of Psychologist-Lawyer Collaboration*, 66 N.C.L. REV. 741, 741 (1988).

³⁹ See McCORMICK, *supra* note 36, at 606-08; MOENSSENS et al., *supra* note 37, at 6-13.

dures can be in widespread use in forensic laboratories, achieving the appearance of general acceptance, yet never be scientifically reliable. Two good examples are the once-popular paraffin test, thought to determine if a suspect had recently fired a gun,⁴⁰ and chirography (handwriting identification).⁴¹ Both have been routinely admitted by courts as generally accepted procedures; neither has a valid scientific basis.

The converse is also true: some scientifically reliable testing procedures have not yet found general acceptance. *Frye* thus produces an inevitable "cultural lag" problem.⁴² Time must pass in order for a new procedure to gain general acceptance, even after the procedure has become scientifically reliable. All new techniques must go through a probationary period, wherein the tests are producing scientifically reliable results, but trials are conducted without them.⁴³ This delay until a procedure becomes generally accepted can be considerable, especially if a new scientific advance is perceived as threatening or too radical (and therefore unacceptable) by established, older scientists in the field.⁴⁴

2. *Frye is Bad Law*

The *Frye* test is also bad law. Its terms are vague and ambiguous, and judges have difficulty applying it to novel scientific evidence.⁴⁵

In order to apply the *Frye* doctrine, a court must first identify the appropriate scientific field. New scientific evidence, however, may not be easily classifiable into a traditional field, or may overlap several fields.⁴⁶ How the court interprets the term "field" may therefore be dispositive. If the court defines the field broadly, it virtually assures that many scientists in that broad field will not have

⁴⁰ See Gianelli, *supra* note 33, at 1224-25; Peter J. Neufeld & Neville Colman, *When Science Takes the Witness Stand*, 262 SC. AM. 46, 49 (May 1990).

⁴¹ See D. Michael Risinger et al., *Exorcism of Ignorance as a Proxy for Rational Knowledge: the Lessons of Handwriting Identification "Expertise"*, 137 U. PA. L. REV. 731 (1989).

⁴² Constantine J. Maletskos and Stephen J. Spielman, *Introduction of New Scientific Methods in Court*, in LAW ENFORCEMENT SCIENCE AND TECHNOLOGY 957, 958 (S.A. Yefsky ed. 1967).

⁴³ See Gianelli, *supra* note 33, at 1223.

⁴⁴ See THOMAS S. KUHN, *THE STRUCTURE OF SCIENTIFIC REVOLUTION* 6-8 (2d ed. 1970); MOENSSEN et al., *supra* note 37, at 6-7.

⁴⁵ If a testing procedure has been around a long time and still not gained general acceptance (e.g., the polygraph), the *Frye* test is indeed one measure of the scientific unreliability of the test. If the testing procedure is new, its lack of general acceptance may be due to either unreliability or novelty. In this case, the methodology has been in general use in speech science for over forty years; only the particular application is new.

⁴⁶ See Petrosinelli, *supra* note 36, at 319.

heard of a new procedure. For example, *People v. King*⁴⁷ involved the admissibility of spectrogram (voiceprint) evidence. The judge defined the relevant field as anatomy, physiology, physics, psychology and linguistics.⁴⁸ The test results were excluded when the proponent could not prove general acceptance in all these disparate disciplines.

There is also danger in defining the field too narrowly. In the case of a dispute within a discipline, those who dissent from the conventional wisdom may form their own subfield, within which untested or unreliable evidence is generally accepted. For example, *People v. Williams*⁴⁹ involved the admissibility of a controversial Nal-line test for detecting narcotic use. The court defined the relevant field as "those who would be expected to be familiar with its use." The test results were therefore easily admitted, although the evidence showed that the medical profession generally had never heard of this supposed test.⁵⁰

Once the field has been defined, the court must decide if a testing procedure has been generally accepted. This language is obviously ambiguous. How many scientists must agree? How unanimous must the agreement be? What is the effect of the opponent's producing "experts" who dispute acceptance? Is any paid consultant's opposing testimony enough to disprove general acceptance? These issues have never been answered by courts interpreting the *Frye* test.⁵¹

Even if general acceptance were more clearly defined, the *Frye* test does not set forth any specific foundation requirements. In the first place, it does not place any restrictions on who may be called as a witness. The general acceptance language could be interpreted as requiring a disinterested foundation witness *other* than a scientist who helped create the new technique.⁵² Such a disinterested witness, however, is unlikely to know enough about the new technique to answer questions (especially on cross-examination) about the details of its operation. Witnesses who know most about the details of a particular test, however, are often not scientists, but technicians or "examiners" who are trained to use a piece of testing equipment, but who are unable to provide necessary information to determine

⁴⁷ 266 Cal. App. 2d 437, 72 Cal. Rptr. 478 (1968).

⁴⁸ *Id.* at 456, 72 Cal. Rptr. at 490.

⁴⁹ 164 Cal. App. 2d Supp. 858, 331 P.2d 251 (Cal. App. Dep't Super. Ct. 1958).

⁵⁰ *Id.* at 860-61, 331 P.2d at 253-54.

⁵¹ See Gianelli, *supra* note 33, at 1215-16.

⁵² *E.g.*, *People v. Tobey*, 401 Mich. 141, 145-46, 257 N.W.2d 537, 539 (1977) (rejecting voiceprint evidence because expert witness built career on voiceprint and was not disinterested).

its underlying scientific reliability.⁵³

It is also unclear whether a *single* witness can establish *general* acceptance. A few courts have held that more than one expert is required,⁵⁴ although this conflicts with the general evidence principle that a foundation may be laid by a single person unless a rule clearly requires corroboration.

Is general acceptance a matter merely of expert *opinion*, or is some minimal level of conditional fact required? For example, may a procedure ever be said to be generally accepted as scientifically valid based on a *single* validating study? The *Frye* test seems to assume that additional validating studies will be done over time — otherwise there is no reason to require a probationary period for new scientific techniques. Some courts, however, have admitted voiceprint evidence on the basis of a single study from Michigan State University.⁵⁵ If additional validating studies are required, how many are required? Must they be done by a second research team? Is publication in a peer-reviewed professional journal sufficient? Is any particular degree of significance required in the validating studies? Must the studies show that a test is 98% reliable? 90%? 75%?

Frye refers to a requirement that some “thing” have gained general acceptance.⁵⁶ It does not explain what that thing is. Although it seems obvious that the court was referring to the specific testing procedure at issue — use of a device that measured systolic blood pressure changes to detect deception — in the same paragraph, the *Frye* court refers to scientific principles, discoveries, and deductions. It leaves open the issue of whether general acceptance must be proved for the underlying scientific theory,⁵⁷ the technology,⁵⁸ the procedure and technique,⁵⁹ the particular instrument used,⁶⁰ or some combination. This can make a substantial difference, because novel scientific evidence may involve a new theory, a new applica-

⁵³ See Gianelli, *supra* note 33, at 1214-15; MOENSSENS et al., *supra* note 37, at 9.

⁵⁴ *E.g.*, *People v. Kelly*, 17 Cal. 3d 24, 37, 549 P.2d 1240, 1248, 130 Cal. Rptr. 144, 157 (1976) (rejecting voiceprint evidence and doubting whether single witness can ever satisfy foundation).

⁵⁵ See Gianelli, *supra* note 33, at 1213.

⁵⁶ *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923).

⁵⁷ See, *e.g.*, *United States v. Addison*, 498 F.2d 741, 743 (D.C. Cir. 1974) (theory must be generally accepted).

⁵⁸ See, *e.g.*, *United States v. Stifel*, 433 F.2d 431, 438 (6th Cir. 1970) (state of technology — neutron activation analysis).

⁵⁹ See *eg.*, *People v. Law*, 40 Cal. App. 3d 69, 84, 114 Cal. Rptr. 708, 718 (1974) (reliability of procedure — voiceprint); *Reed v. State*, 283 Md. 374, 385, 391 A.2d 364, 370-71 (1978) (reliability of technique or process — voiceprint).

⁶⁰ See *Commonwealth v. Fatalo*, 346 Mass. 266, 269, 191 N.E.2d 479, 481 (1963) (scientific acceptance of instrument — polygraph).

tion of established theory, an improved procedure, or the use of a new instrument.⁶¹

Finally, *Frye* is bad law because it requires judges to make decisions for which they are ill-equipped. Judges generally are illiterate in science, untrained in statistics, and operate in a legal culture that is non-scientific (if not actively hostile to science).⁶² Judges are therefore poorly equipped to distinguish generally reliable from unreliable methods. Nor are they likely to get any help from lawyers, who are similarly untrained in science and have difficulty accessing scientific debates in science journals.⁶³

3. *Frye Is Premised on Bad Psychology*

One underlying premise of the *Frye* test is that jurors will be easily awed by expert testimony. Gianelli asserts that "an aura of scientific infallibility may shroud the evidence and thus lead the jury to accept it without critical scrutiny."⁶⁴ Professor Laurence Tribe makes the same argument about statistical evidence.⁶⁵ In *U.S. v. Addison*, the D.C. Court of Appeals asserts that scientific evidence may "assume a posture of mystic infallibility in the eyes of a jury of laymen."⁶⁶

This assertion about human behavior is dubious. Social psychologists have demonstrated that people generally *undervalue* scientific data, misunderstand and under-utilize statistics, rely on anecdotes and emotion rather than empirical scientific evidence when making important decisions, and persistently hold beliefs contrary to scientific logic and mathematics.⁶⁷ Some studies focusing specifically on juries have found that expert witnesses have no sig-

⁶¹ See Gianelli, *supra* note 33, at 1212.

⁶² J. Alexander Tanford, *The Limits of a Scientific Jurisprudence: the Supreme Court and Psychology*, 66 IND. L.J. 137, 154-55 (1990).

⁶³ See Neufeld & Colman, *supra* note 40, at 49.

⁶⁴ Gianelli, *supra* note 33, at 1237.

⁶⁵ Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1355 (1971).

⁶⁶ 498 F.2d 741, 744 (D.C. Cir. 1974). See also *State v. Carlson*, 267 N.W.2d 170, 176 (Minn. 1978).

⁶⁷ See RICHARD E. NISBETT & LEE ROSS, HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT 55-56 (1980); Michael Saks & Robert Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 LAW & SOC'Y REV. 123, 127-31, 149 (1981); William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy*, 11 LAW & HUM. BEHAV. 167 (1987); William C. Thompson, *Are Juries Competent to Evaluate Statistical Evidence*, LAW & CONTEMP. PROBS., Autumn 1989, at 9, 33 (most subjects give less weight to statistical evidence than it deserves).

nificant impact on verdicts.⁶⁸

B. ALTERNATIVE LEGAL STANDARDS

In response to the weaknesses of the *Frye* test, commentators have suggested that it be replaced with either of two alternatives: A modified *Frye* test or a relevancy test.⁶⁹

1. Modified Frye Tests

One proposal would modify the "general acceptance" part of the *Frye* test. Commentators have suggested the substitution of "substantial" or "reasonable" acceptance.⁷⁰ This would probably have the effect of admitting more scientific evidence, but would still be ambiguous and difficult to apply. It also would be unlikely to more effectively distinguish reliable from unreliable *new* techniques.

A second proposal would modify the definition of "field." In *People v. Williams*,⁷¹ the court stated that it would accept scientific evidence unknown in a general scientific field, if it were accepted as reliable within a narrow specialty—those who would be expected to be familiar with its use. This variation would naturally tend to admit more novel scientific evidence, but it fails to address the problem that even within a specialty field, acceptance is not the same thing as reliability.

The third proposal is to modify *Frye's* requirement that the technique be generally accepted. It would require only that the scientific principles underlying a new testing procedure be generally accepted. New testing procedures designed to explore a particular problem using generally accepted principles and existing equipment would be admissible, even if the particular application were new.⁷²

⁶⁸ E.g., Martha A. Myers, *Rule Departures and Making Law: Juries and Their Verdicts*, 13 LAW & SOC'Y REV. 751, 762 (1979).

⁶⁹ A third alternative has also been suggested: delegating the decision on reliability to independent bodies of science advisors. Some have suggested the creation of a science court, others the use of independent bodies of experts hired to advise courts on validity of scientific techniques. See Gianelli, *supra* note 33, at 1231-32; McCORMICK, *supra* note 36, at 608 n.27. The idea was incompatible with the adversary system, and went nowhere. See David L. Bazelon, *Coping with Technology Through the Legal Process*, 62 CORNELL L. REV. 817, 826-28 (1977).

⁷⁰ JAMES R. RICHARDSON, MODERN SCIENTIFIC EVIDENCE 24 (2d ed. 1974) (substantial acceptance); Howard A. Latin et al., *Remote Sensing Evidence and Environmental Law*, 64 CAL. L. REV. 1300, 1380 (1976) (reasonable acceptance); Lucinda E. Minton, Note, *Expert Testimony Based on Novel Scientific Techniques: Admissibility Under the Federal Rules of Evidence*, 48 GEO. WASH. L. REV. 774, 787 (1980) (preponderance of experts accept).

⁷¹ 331 P.2d 251 (1958).

⁷² *Ibn-Tamas v. United States*, 407 A.2d 626, 638 (D.C. 1979); *Coppolino v. State*, 223 So. 2d 68 (Fla. Dist. Ct. App. 1968), *appeal dismissed*, 234 So. 2d 120 (Fla. 1969), cert. denied, 399 U.S. 927 (1970). Gianelli argues that the *Coppolino* test is essentially the

2. Relevancy Tests

The trend, however, is to reject *Frye* rather than to attempt to modify it. As early as 1954, leading scholars were calling for the substitution of a basic relevancy test. Any relevant scientific evidence supported by a qualified expert should be presumptively admissible.⁷³ Neither the evidence's novelty nor its lack of general acceptance are dispositive, for neither criterion makes scientific evidence *relevant*.⁷⁴ The primary check against unreliable or pseudo-scientific evidence is the rule permitting the opponent to contradict the expert's testimony by introducing passages from leading textbooks in the field written by reliable authorities.⁷⁵

Scholars who advocate the use of a basic relevancy test have suggested five criteria which scientific evidence sought to be admitted should meet. First, the evidence must be introduced through a properly qualified expert. Second, the subject matter must be one on which expert testimony will assist the jury. Third, all equipment involved in generating the evidence must be shown to be in good working order. Fourth, the evidence must be relevant under basic relevancy rules. Fifth, the probative value of the evidence must be "weighed" against its potential prejudicial effect. Each of these criteria will be discussed in turn.

a. Introduction by an Expert Witness

The first admissibility requirement under a relevancy approach is that *scientific evidence must be introduced through a properly qualified expert*. All novel scientific evidence must be sponsored by an expert witness who can explain both the theoretical and practical reliability of the new testing procedure. Although the usual rule is that either education or experience will qualify a person as an expert,⁷⁶ in the Hazelwood case a fully-educated scientist is probably required. Mere experience with a new forensic technique is inadequate to explain the theoretical validity of a procedure, so a technician probably cannot lay the necessary foundation.⁷⁷

same thing as McCormick's relevancy test. Gianelli, *supra* note 33, at 1233-35. See text accompanying notes 73-91, *infra*.

⁷³ 1 CHARLES T. MCCORMICK, EVIDENCE 363-64 (1954).

⁷⁴ *United States v. Stifel*, 433 F.2d 431, 438 (6th Cir. 1970) (newness and lack of absolute certainty affect weight, not admissibility: "Every useful new development must have its first day in court.").

⁷⁵ See FED. R. EVID. 803(18).

⁷⁶ *E.g.*, FED. R. EVID. 702.

⁷⁷ See Gianelli, *supra* note 33, at 1214-15; MOENSSENS et al., *supra* note 37, at 9-10.

b. Expert Testimony Must Assist the Jury

The second requirement is that *the subject matter must be one in which expert testimony will assist the jury*. Expert testimony generally is permitted only in situations in which the jury could use some help. For example, experts with breathalyzers and blood test results will be of real assistance to a jury in determining a person's blood-alcohol content, but traffic safety experts are probably not needed to help a jury determine if high-speed drunk driving is dangerous. The difference is whether the jury is presumed to be competent to draw their own conclusions based on observations.

There are three possible situations in which introduction of expert testimony might be sought. First, the data upon which conclusions are to be drawn may be beyond the jurors' perceptions, so that jurors are incapable of drawing any conclusions about the subject, e.g., quantum mechanics or brain surgery. In this case, experts will always be permitted. Second, the data may consist of common subjective evaluations based on perceptions completely within common knowledge, e.g., whether a concrete block falls to the ground when dropped off a scaffold. No physicist is necessary to explain gravity. In this situation, experts are superfluous. Third, the subject may be somewhere in between. The data may be partially hidden, or it may be an area in which lay and expert witnesses use quite different methods for reaching similar kinds of conclusions, e.g., whether failure to water a lawn during a drought caused grass to die.

Under traditional principles of evidence law, still followed in a few states, scientific evidence was presumed inadmissible. Expert testimony was permitted only in the first situation — when the matter at issue was completely beyond the understanding and common experience of the average juror.⁷⁸ Thus, scientific evidence was excluded if the jurors and the expert were similarly qualified to draw conclusions. The rule did not recognize the possibility that jurors might draw *better* conclusions with the aid of experts.

The modern version of the rule⁷⁹ reflects the contemporary view of presumptive admissibility. It states that scientific evidence is admissible if it will “assist” the jury in understanding the evidence or in drawing conclusions. Under this version, evidence should be *admitted* if jurors and experts can *both* draw reliable conclusions. Indeed, it is under this modern view that testimony on field sobriety

⁷⁸ E.g., *State v. Maudlin*, 416 N.E.2d 477 (Ind. Ct. App. 1981). See McCORMICK, *supra* note 36, at 33.

⁷⁹ E.g., FED. R. EVID. 702.

tests is admitted.⁸⁰

c. Equipment in Good Working Order

The third requirement is that *all equipment involved in generating scientific evidence be shown to be in good working order*. Inherent in the creation of scientific evidence is reliance on machinery, technology, and laboratory equipment. The reliability of an ultimate scientific conclusion may depend on factors the scientist takes for granted — that ordinary pieces of apparatus were in good working order and were operated by qualified persons. The law requires proof that instruments and equipment, such as microphones, tape recorders, and X-ray machines, were in good working order and were properly operated during the creation of any particular piece of scientific evidence. It also requires that human beings be accounted for. If an expert relies on lab technicians or graduate students, they must be shown to be reliable and properly trained. Some jurisdictions also require proof that reliance on these supporting procedures be considered reasonable by experts in the field.⁸¹

d. Scientific Reliability (Relevance)

The fourth requirement is that *scientific evidence must be relevant under basic relevancy rules*, such as Federal Rules of Evidence (FRE) 401-402. Under FRE 401, evidence is relevant whenever it helps the jury determine the facts in issue to any extent. Gianelli argues that scientific evidence will help the jury when it can be shown to be reasonably *reliable* under a three part test: 1) The underlying principles must be considered valid by the scientific community. The test must be based on conventional scientific theory. 2) The technique applying the principle must be scientifically reliable. Has basic research been conducted that demonstrates that the procedure works as predicted, and generally produces statistically significant results? 3) The technique must have been applied properly on the occasion of the particular test.⁸² Novelty and lack of general acceptance do not negate reliability, and thus go to the weight rather than the admissibility of the evidence.⁸³

⁸⁰ See *People v. Randolph*, 213 Cal. App. 3d Supp. 1, 262 Cal. Rptr. 378 (1989); *People v. Krueger*, 99 Ill. App. 2d 431, 437-38, 241 N.E.2d 707, 712 (1968). Typical field sobriety tests are described in *Pennsylvania v. Muniz*, 110 S. Ct. 2638, 2641 n.1 (1990).

⁸¹ *E.g.*, FED. R. EVID. 703.

⁸² Gianelli, *supra* note 33, at 1201-02. See also Neufeld & Colman, *supra* note 40, at 48 (three-part test is sensible from scientific point of view).

⁸³ Degree of acceptance, including testimony by opposing experts, goes to weight,

This relevancy test creates a standard of presumptive admissibility. A judge "should exclude an expert opinion *only* if it is 'fundamentally unsupported' and 'would not actually assist the jury in arriving at an intelligent and sound verdict.'"⁸⁴ As a rule, "it is better to admit relevant scientific evidence in the same manner as other expert testimony and allow its weight to be attacked by cross-examination and refutation."⁸⁵ Thus, any relevant conclusions supported by even a *single* qualified expert witness should be received unless there are distinct reasons for excluding them.⁸⁶

e. Balancing Probative Value Against Prejudicial Effect

The fifth requirement is that *the probative value of the scientific evidence must be "weighed" against its potential prejudicial effect.*⁸⁷ Under the federal rule, the evidence will be admitted unless its probative value is *substantially* outweighed by the danger that it will mislead the jury.⁸⁸ If the evidence helps to prove one of the central disputed issues in a case, it will almost always be admissible.⁸⁹ The mere fact that evidence is scientific in nature does not make it prejudicial.

Courts have generally articulated only one major danger: scientific evidence may mislead a jury into making a *factual* error because of a supposed aura of scientific infallibility.⁹⁰ Social

not admissibility. *Jenkins v. State*, 156 Ga. App. 387, 388, 274 S.E.2d 618, 619 (1980); *Reed v. State*, 283 Md. 374, 386-87; 391 A.2d 364, 370-71 (1978).

⁸⁴ *Christophersen v. Allied Signal Corp.*, 902 F.2d 362, 364-65 (5th Cir. 1990)(emphasis added)(quoting *Viterbo v. Dow Chemical Co.*, 826 F.2d 420, 422 (5th Cir. 1987)). *But see Gianelli, supra* note 33, at 1245-50 (proposing a presumption of inadmissibility and a requirement that the state prove reliability beyond a reasonable doubt).

⁸⁵ *United States v. Baller*, 519 F.2d 463, 466 (4th Cir. 1975).

⁸⁶ *Cf. Gianelli, supra* note 33, at 1236, 1243 (criticizing test for this reason). McCormick argues that one expert is not enough. He would require some corroboration, such as publication in a peer-reviewed journal, and that scientists rather than technicians serve as expert witnesses. *McCORMICK, supra* note 36, at 609.

⁸⁷ *See J. Alexander Tanford, A Political-Choice Approach to Limiting Prejudicial Evidence*, 64 *IND. L.J.* 831, 859-71 (1989) (criticizing balancing metaphor and suggesting that courts are really making choices).

⁸⁸ *See Gianelli, supra* note 33, at 1230. *FED. R. EVID.* 403 also provides for the exclusion of relevant evidence that will cause undue prejudice, confusion of the issues, and undue waste of time. Scientific evidence does not substantially implicate any of these. Unfair prejudice refers to the arousing of emotions in jurors, when those emotions are not inherent in the nature of the case. *Tanford, supra* note 87, at 843. Confusion of the issues generally refers to confusion about the *legal* issues. If evidence tends to cause jurors to apply an erroneous understanding of the law, it is prejudicial. *Id.* at 847-48. Undue waste of time usually means that undue time would be spent on *tangential issues*, and does not apply to evidence that casts light on the central issues. *Id.* at 852-54.

⁸⁹ *See Tanford, supra* note 87, at 863-64 (if evidence has any real probative value, courts admit it).

⁹⁰ *See Gianelli, supra* note 33, at 1237. If techniques are demonstrable in the courtroom and involve principles and procedures understandable to the lay jury, there is little

psychologists, however, suggest that the risk of undue influence is minimal — certainly far short of the “substantially outweighs” standard of Rule 403. The better test would find substantial danger of misleading the jury only when “an exaggerated popular opinion of the accuracy of a particular technique makes its use . . . likely to mislead the jury.”⁹¹

A supposedly different version of the relevancy test was suggested in *United States v. Downing*.⁹² The *Downing* test bases the admissibility of scientific evidence on Rule 702 rather than Rules 401-403.⁹³ It assumes that the use of the word “assist” in Rule 702 means more than mere relevancy. The *Downing* court suggested a balancing analysis that asks: 1) How reliable is the process or technique used in generating the evidence? 2) Will it overwhelm, confuse or mislead the jury? and 3) How strong is the connection between the test results and the factual dispute? This test seems virtually indistinguishable from the basic relevancy test.

A few appellate courts have endorsed a more informal, ad hoc determination of reliability.⁹⁴ These courts apparently focus primarily on the third criteria of the relevancy test: was the test properly conducted in the particular case? This is clearly a question of fact for the trial judge. The problem with this approach is that it grants to thousands of trial judges the power to make individual rulings on the validity of scientific theories and techniques. Trial judges, however, will not all be equally scientifically literate, and their rulings are likely to vary. Yet, the very essence of science is its universality.⁹⁵ It does not vary from county to county and case to case, so a rule of law permitting such results is unsound.⁹⁶

concern with experts having undue influence. However, when the nature of the analysis is esoteric or invisible, e.g., DNA-typing that depends on knowledge of molecular biology, chemistry, genetics and statistics, a stronger showing of probative value could be required. See William C. Thompson & Simon Ford, *DNA Typing: Acceptance and Weight of the New Genetic Identification Tests*, 75 VA. L. REV. 45, 52 (1989).

⁹¹ *United States v. Baller*, 519 F.2d 463, 466 (4th Cir. 1975). See Andrew E. Taslitz, *Does the Cold Nose Know? The Unscientific Myth of the Dog Scent Line-up*, 42 HASTINGS L.J. 15, 23-28, 42-50 (1990) (popular belief in infallibility of bloodhounds' powers of scent far exceeds scientific fact).

⁹² 735 F.2d 1224 (3d Cir 1985).

⁹³ FED. R. EVID. 702 provides: “If scientific . . . knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, [an] expert . . . may testify thereto in the form of an opinion or otherwise.”

⁹⁴ See *State v. Hall*, 297 N.W.2d 80 (Iowa 1980); *United States v. Stifel*, 433 F.2d 431, 437-41 (6th Cir. 1970).

⁹⁵ ROBERT K. MERTON, *THE SOCIOLOGY OF SCIENCE* 270-73 (Norman W. Storer ed., 1973).

⁹⁶ See Gianelli, *supra* note 33, at 1222-23.

IV. IS ACOUSTIC-PHONETIC EVIDENCE ADMISSIBLE IN THE EXXON VALDEZ CASE?

A. ADMISSIBILITY UNDER THE *FRYE* TEST

We concede at the outset that the results of our analyses of Captain Hazelwood's speech are probably not admissible under the *Frye* test. Although our evidence derives from generally accepted scientific theory, and uses standard techniques of speech analysis, the particular application to alcohol has not yet achieved general acceptance. Despite all the problems of interpreting the meaning of the old *Frye* test, it is doubtful that the evidence would be admitted in any jurisdiction that strictly follows the old rule, which favors exclusion.

It is also unlikely that our evidence would be admissible under two of the modified *Frye* tests. The application probably has not yet gained "substantial" acceptance. It is too new for many speech scientists to have heard of and thought about it. Articles describing the procedure are just now appearing in peer-reviewed scientific journals. Nor has a subspecialty emerged that uses this novel application to measure the effects of alcohol on speech, the existence of which would justify using the "narrow field" test.

This evidence might, however, be admissible under the third *Frye* modification, which requires a new forensic application to be derived from accepted techniques, based on generally accepted underlying scientific principles and using generally accepted laboratory equipment. To lay a foundation under this *Frye* variation, the proponents must show that the underlying scientific theory is considered valid (generally accepted) by the scientific community, and that the techniques and equipment used are known to be reliable and are in widespread use (generally accepted) in the scientific community. These are essentially the same requirements contained in part four of the relevancy test, and will be considered below.

B. ADMISSIBILITY UNDER A RELEVANCY TEST

Whether evidence of acoustic analyses of Captain Hazelwood's speech using audio tapes from the *Exxon Valdez* should be admissible under the modern relevancy test depends on the answers to five questions: 1) Is a properly qualified expert witness available who can explain the theoretical and practical reliability of the tests? 2) Is the subject-matter one on which expert testimony will assist the jury in understanding the evidence and drawing accurate conclusions? 3) Were all pieces of equipment used in the test in good working order, and all technical personnel adequately trained?

4) Are the underlying theories reliable, are the techniques valid, and were the proper procedures followed? 5) Is there an exaggerated popular opinion of the accuracy of the technique so that its use is likely to mislead the jury?

1. *Introduction by an Expert Witness*

Properly qualified experts in speech science must be available to sponsor the evidence. Speech science is a diverse field that draws upon linguistics, psychology, clinical speech and hearing science, physiology, physics and electrical engineering.⁹⁷ The tie that binds the field together is the Acoustical Society of America, which includes a speech communication section. A person with a graduate degree in one of the underlying disciplines, with training or experience focusing on speech science, affiliated with a university- or industry-sponsored speech laboratory, who is a member of the ASA, and who is familiar with the literature on alcohol and speech,⁹⁸ ought to qualify as an appropriate witness.

Because the evidence in the Hazelwood case concerns the effects of alcohol on speech, some experience in the field of alcohol research also is desirable. Even among speech scientists, however, few will have the necessary expertise. The effects of alcohol on speech production is tangential to the main body of acoustic-phonetic research.⁹⁹ Few speech scientists have personally conducted research in the area because the research protocols¹⁰⁰ are more complicated than for usual speech research, and because no obvious theoretical issues are involved. However, at least a few qualified experts are available.¹⁰¹

⁹⁷ See generally BORDEN & HARRIS, *supra* note 9 (basic textbook giving overview of speech science). See also FLANAGAN, *supra* note 18 (textbook on engineering and physics of speech); DENNIS B. FRY, *THE PHYSICS OF SPEECH* (1979) (textbook in speech acoustics); PHILIP LIEBERMAN & SHEILA E. BLUMSTEIN, *SPEECH PHYSIOLOGY, SPEECH PERCEPTION, AND ACOUSTIC PHONETICS* (1988) (textbook on physiology and psychology of speech).

⁹⁸ See, e.g., Johnson et al., *supra* note 1, at 3-11 (summarizing relevant literature).

⁹⁹ See, e.g., BORDEN & HARRIS, *supra* note 9, at 291-302 (detailed index contains no entries relating to research on the effects of alcohol); HARRY HOLLIEN, *THE ACOUSTICS OF CRIME ix-xiv* (1990) (table of contents of recent textbook on acoustics makes no reference to alcohol effects).

¹⁰⁰ Such as obtaining approval from human subject committees.

¹⁰¹ E.g., the second and third authors of this paper. Dr. Pisoni has a Ph.D. in Cognitive Psychology, is a Professor of Psychology and Cognitive Science at Indiana University and currently the Director of its Speech Research Laboratory, has personally conducted research on the effects of alcohol on speech, e.g., Pisoni & Martin, *supra* note 1, and is familiar with the literature. Johnson et al., *supra* note 1, at 216-220. Dr. Johnson has a Ph.D. in Linguistics from Ohio State University, is currently a post-doctoral fellow in the Phonetics Laboratory of the Department of Linguistics at U.C.L.A., has

2. Expert Testimony Will Assist the Jury

If the results of acoustic analyses of Captain Hazelwood's speech are to be admissible, the evidence must assist the jury in determining whether Hazelwood was intoxicated. People draw conclusions about intoxication all the time,¹⁰² so this is not a situation in which experts are required because the data necessary to draw reliable conclusions are inaccessible to the average person. Neither is this a topic considered so completely within common knowledge that experts are never permitted. Courts routinely admit other forms of expert testimony concerning alcohol impairment, such as the results of breathalyzer and blood tests¹⁰³ and field sobriety tests.¹⁰⁴ Therefore, our acoustic-phonetic analyses should be evaluated under the middle category in which particular experts will be allowed if their testimony will assist the jury in drawing accurate conclusions.

Empirical research suggests that jurors often can tell whether a speaker is intoxicated. Many phonetic changes that accompany alcohol impairment, such as word substitutions and revisions, are easily detected by naive listeners. Pisoni and Martin found that both college students and Indiana State Troopers could identify intoxicated (0.10% BAL) talkers with about 80% accuracy when they directly compared intoxicated to sober speech samples.¹⁰⁵ Even when a speech sample was presented in isolation, Martin and Yuchtman found that naive listeners were able to identify intoxicated speakers based on listening with about 66% accuracy. Although these detection rates are better than chance, they also show that jurors can make mistakes. Can expert testimony by speech scientists assist the jury in minimizing these mistakes? The issue is not whether experts can *more accurately* determine whether a speaker has been drinking, but whether they can provide additional data that will assist the jury's common sense determination.

In the Exxon Valdez case, the information that can be provided by experts is nonredundant. Although speech scientists' observations of gross phonetic changes overlap the aural information jurors

personally conducted research on the effects of alcohol on speech, and is familiar with the literature. *Id.*

¹⁰² *E.g.*, *New v. State*, 254 Ind. 307, 259 N.E.2d 696 (1970) (lay people commonly make conclusions about intoxication).

¹⁰³ *E.g.*, *Shuman v. State*, 489 N.E.2d 126 (Ind. Ct. App. 1986) (blood serum tests for intoxication admissible); *Ballou v. Henri Studios*, 656 F.2d 1147 (5th Cir 1981) (breath tests admissible).

¹⁰⁴ *See* *People v. Randolph*, 213 Cal. App. 3d Supp. 1, 262 Cal. Rptr. 378 (1989); *People v. Krueger*, 99 Ill. App. 2d 431, 241 N.E.2d 707 (1968).

¹⁰⁵ Pisoni & Martin, *supra* note 1.

rely on, their instrumental measurements provide data otherwise unavailable to jurors — objective, unbiased quantitative measurements of segmental and suprasegmental effects, such as pitch, duration, and amplitude.¹⁰⁶ Like breathalyzer measurements, expert evidence based on acoustic-phonetic analyses of speech would provide *unbiased* data to supplement jurors' intuitions. This would assist, rather than hinder, the jury in reaching a more accurate conclusion about whether a speaker was intoxicated.

3. *Equipment in Good Working Order*

To satisfy the third part of the foundation for scientific evidence, all equipment involved in acoustic analyses of speech samples must have been in proper working order at all times in question. This requirement would be an essential part of speech science research¹⁰⁷ even if it were not part of the legal foundation. One measure of the true scientific nature of tests and experiments is the degree of care taken to make sure all the equipment was in good working order and operated properly by trained technicians.

Because the analysis procedures used on the *Exxon Valdez* tapes involved a comparison across speech samples, rather than absolute detection of impairment, the only real requirement for the recordings is that the same recording and playback equipment was used for both samples in the comparison. Most imperfections in equipment would appear on both the control and test recordings and would not affect analysis. For example, if a tape recorder runs slightly slowly, the resulting tapes when played back will show increased average pitch compared to the person's *true* pitch. We are concerned, however, only with *relative* changes in pitch between samples, a factor unaffected by the overall change in tape speed.

Concededly, a tape recorder could suffer from wow and flutter or other erratic speed fluctuations that could affect tape-to-tape comparisons. The possibility of unusual speed variations in this case was minimized by verifying that the speaking rate of a presumably sober Coast Guard radio operator on the same tapes did not vary significantly. In addition, we analyzed background noise across tape samples and found that its frequency did not vary significantly. These two sets of measurements strongly indicate that the equipment was properly functioning to the extent necessary for drawing reliable conclusions based on tape-to-tape comparisons.

In order to minimize the possibility of error in the laboratory, it

¹⁰⁶ See *supra* Tables 3-6.

¹⁰⁷ See BORDEN & HARRIS, *supra* note 9, at 224-25.

is common to take multiple measurements using different personnel. We followed such a procedure in analyzing the tapes of Captain Hazelwood's speech. Two well-trained post-graduate researchers made independent, overlapping measurements. No errors were detected when we compared them. This is standard operating procedure in speech science laboratories around the country.¹⁰⁸

4. *Scientific Reliability and Relevance*

The evidence concerning Captain Hazelwood's possible alcohol impairment is sufficiently reliable to be relevant under FRE 401 and 402.

a. Theory

The general principles of speech science are considered scientifically valid and are not controversial.¹⁰⁹ The specific theoretical assumption underlying our acoustic analyses of Captain Hazelwood's speech samples is that alcohol affects motor function which affects speech control.¹¹⁰ This, too, seems largely undebateable. After all, it is one of the major assumptions underlying field sobriety tests routinely used by law enforcement agencies as preliminary evidence of alcohol impairment. Another indication of the validity of this theory is the publication in peer-reviewed scientific journals of three articles describing this research and its theoretical underpinnings.¹¹¹ No articles have been published that suggest any contrary view. Indeed, the results are consistent with the findings from over forty years of basic research on speech acoustics and speech production.¹¹²

b. Validation of Technique

The general techniques of acoustic-phonetics that we employed are in common use in speech laboratories throughout the country. Previous research has validated them as reliable methods of measur-

¹⁰⁸ See *id.*, at 223-53 (describing standard acoustic-phonetic research protocols).

¹⁰⁹ Speech science generally is described in BORDEN & HARRIS, *supra* note 9, and FLANAGAN, *supra* note 18. Acoustic-phonetic analysis of speech sounds is described in BORDEN & HARRIS, *supra*, at 223-53, and by FLANAGAN, *supra*, at 140-204.

¹¹⁰ See Berry & Pentreath, *supra* note 7, at 43-72; Johnson et al., *supra* note 1, at 216-17; Klingholz et al., *supra* note 15, at 929-35.

¹¹¹ See Pisoni & Martin, *supra* note 1, at 577-78; Johnson et al., *supra* note 1, at 217-20; Klingholz et al., *supra* note 15.

¹¹² See BORDEN & HARRIS, *supra* note 9, at 131-34, 142-50 (theoretical models of speech production); Flanagan, *supra* note 18, at 9-23 (physiology of speech production). See also *id.* at 45-58, 78-89, 106-117 (basic description of how speech is produced).

ing speech effects.¹¹³

Previous controlled studies of subjects other than Captain Hazelwood demonstrate that these techniques can reliably measure when a person's speech has been affected by alcohol consumption. Acoustic analyses cannot *prove* that a person was definitely intoxicated, nor specify the exact blood alcohol level. However, the Pisoni and Martin study revealed that alcohol affects speech in predictable ways, producing patterns of measureable effects more consistent with significant alcohol impairment than with any other known condition that affects speech.¹¹⁴ The results of the Pisoni and Martin validating study are further corroborated by other experimental studies of alcohol-related speech impairments.¹¹⁵

5. *Balancing Probative Value Against Prejudicial Effect*

Because the scientific evidence concerning Captain Hazelwood's speech is reliable, it should be admitted unless some Rule 403 danger *substantially* outweighs its probative value. Assuming the evidence is introduced in a case in which the possible intoxication of Captain Hazelwood is an important issue, its probative value is high. It is the only evidence of its kind. The duration and frequency measurements are unique because they are the only physical evidence and the only *unbiased* evidence bearing on the issue of intoxication. The only other available evidence consists of the opinions of eyewitnesses who worked for Exxon. Nonredundant evidence on a central issue will almost always be admissible,¹¹⁶ because it would require extreme Rule 403 dangers to significantly outweigh high probative value.

There is no problem here with the usual danger associated with scientific evidence — misleading the jury into making a *factual* error because of an exaggerated popular opinion of the accuracy of a particular technique. There is little, if any, evidence that the public has ever heard of acoustic-phonetic analysis of speech.

Are there any other possible FRE 403 dangers? The only other likely objection to it is that it may constitute an undue waste of time. This objection is normally unavailing when the evidence goes to the heart of an issue, as this evidence does. It is unlikely to be consid-

¹¹³ *Eg.*, BORDEN & HARRIS, *supra* note 9, at 224-28, 230-36, 250-53; Flanagan, *supra* note 18, at 141-55, 184-86.

¹¹⁴ Pisoni & Martin, *supra* note 1.

¹¹⁵ *See* Klingholz et al., *supra* note 15; Lester & Skousen, *supra* note 10; Pisoni et al., *supra* note 10; Sobell & Sobell, *supra* note 8; Sobell et al., *supra* note 8; Trojan & Kryspin-Exner, *supra* note 10.

¹¹⁶ *See* Tanford, *supra* note 87, at 863-70.

ered an *undue* waste of time to allow a battle of experts on the central issue—whether Hazelwood was drunk. In any event, battles occur mostly when scientific evidence requires considerable subjective interpretation based on the absence of hard data — psychiatric diagnosis is the paradigmatic example. Our conclusion that Hazelwood's voice showed effects consistent with alcohol consumption is based in part on objective data (i.e., physical measurements of duration and fundamental frequency) requiring little interpretation. Nothing indicates the likelihood of any real debate over the conclusions that we draw. The opposition will most likely take the form of objections to the methodology employed. That battle will be fought mostly in front of the judge on the question of admissibility. Therefore, there is no serious risk of prejudice that can *substantially* outweigh the relevancy of the evidence.

V. CONCLUSIONS

Expert testimony, based on acoustic analyses of audio tapes, that Captain Hazelwood probably was intoxicated at the time the *Exxon Valdez* ran aground should be admitted. Properly qualified experts are available to sponsor the evidence. The evidence will assist the jury in determining one of the key facts in issue. The analyses of the Hazelwood tapes appear to have been conducted properly. The evidence itself is scientifically reliable. It is based on accepted theories of speech acoustics and uses standard equipment and technology. The accuracy of these techniques has already been demonstrated in controlled laboratory experiments in which subjects were intoxicated to known blood alcohol levels. Finally, no particular fact-finding danger is posed by its use. Accordingly, these analyses should be admissible under the emerging "relevancy test" for scientific evidence.