Issues in Forensic Voice

***Harry Hollien, †Ruth Huntley Bahr, and *James D. Harnsberger,** **Gainesville and †Tampa, Florida*

Summary: The following article provides a general review of an area that can be referred to as Forensic Voice. Its goals will be outlined and that discussion will be followed by a description of its major elements. Considered are (1) the processing and analysis of spoken utterances, (2) distorted speech, (3) enhancement of speech intelligibility (re: surveillance and other recordings), (4) transcripts, (5) authentication of recordings, (6) speaker identification, and (7) the detection of deception, intoxication, and emotions in speech. Stress in speech and the psychological stress evaluation systems (that some individuals attempt to use as lie detectors) also will be considered. Points of entry will be suggested for individuals with the kinds of backgrounds possessed by professionals already working in the voice area.
Key Words: Forensic–Intelligibility–Surveillance–Recordings–Speech–Evaluation Systems–Identification–Psychoacoustics–Linguistics–Phonetics.

INTRODUCTION

This essay is based on the special program that was proffered at the 2012 Symposium, Care of the Professional Voice.[1] The major topics presented there also will be reviewed here. As with the oral presentations, they will be directed to those professionals working within one or more of the domains encompassed by what can be referred to as "Voice."

As you would expect, we will attempt to be sensitive to the differences among and between the members of the several subdisciplines found there. They can be identified as voice scientists (often phoneticians), medical practitioners, speech-language pathologists, voice teachers, and related musicians. It is recognized that their skills, experience, and training may vary but the commonalities among them are rather substantial. More importantly, professionals in each of these areas can adapt their basic expertise for use in the specialized area of Forensic Voice.

Of course, the scientists and medical practitioners probably could make a reasonably meaningful transfer simply by appropriately modifying their procedures/techniques, learning about certain new ones, and becoming familiar with the needs of law enforcement, judicial, and intelligence agencies. Moreover, because individuals working in the clinical areas of voice/speech have substantial training (and experience) both in the productions and perception of voice, speech, and language, they too have reasonable basic preparation. Finally, although individuals working in music and singing may not have as much technical knowledge as the specialists listed above, their training in music and languages—as well as their ability to perceive and detect minute variations in heard signals—would serve them well in many Forensic Voice areas. Indeed, it is clear that this fourth group may not yet appreciate the level of their auditory skills for detecting and processing signals that are distorted, masked, or incomplete.

In this regard, it would appear necessary to provide some evidence as to just how sensitive and sophisticated the human auditory mechanism is for all four groups and how effectively it can be used to process acoustic and related neural signals. Specifically, Kraus and Nichol[2] provide a general review of some of the auditory differences found between musicians and control subjects. They describe several of the absolutely remarkable auditory processing abilities possessed by *both* groups—and, hence, by humans in general. In doing so, they (and others) describe how well voice embedded in noise can be processed,[3,4] the speed by which complex sounds can be decoded (ie, in only 10–12 milliseconds), the enhancement of such processing by training and/or experience,[5–7] how emotions can be coded,[6] how language (including tone languages), plus speech and speakers, can be assessed,[5,8,9] and so on. Other research in this area supports the validity of their positions[8,10] as does that of DeJong[11] in her research on earwitness identification. In short, the hypothesis that the auditory system is capable of substantial finite processing of the type necessary for accuracy in several subareas within Forensic Voice cannot be rejected (see also certain of the experiments on speaker identification [SI] reported in that section). In short, what is needed for all groups is some technical processing information and an understanding of the needs of the Forensic Sciences.

Definitions

Forensic Voice may be considered to be part of Forensic Communication, which, in turn, is situated within the Forensic Sciences. It is made up of three related subdisciplines, all of which are structured to meet the relevant needs of Criminal Justice, Judicial, and Intelligence agencies. The three areas are:

1) Forensic Linguistics, including Psycholinguistics, targets language (written or spoken) which is analyzed to determine authorship, the intent of the individual, deception, and so on. The area also includes speech/language decoding, a task shared with Forensic Voice.
2) Forensic Psychoacoustics relates to human hearing or audition. In this case, the analyses involve heard signals and their effect (acoustic, perceptual, and neural) on individuals and their behaviors.
3) Forensic Voice (most often referred to as Forensic Phonetics[12]) focuses on the analysis of spoken communication for the purposes cited above. It includes enhancing and decoding spoken messages, SI, analysis of emotions

Accepted for publication June 13, 2013.
From the *Institute for Advanced Study of the Communicative Processes (IASCP), University of Florida, Gainesville, Florida; and the †Communication Sciences and Disorders, University of South Florida, Tampa, Florida.

Address correspondence and reprint requests to Harry Hollien, MD, 229 SW 43rd Terrace, Gainesville, FL 32607. E-mail: hollien@ufl.edu
Journal of Voice, Vol. 28, No. 2, pp. 170-184
0892-1997/$36.00
© 2014 The Voice Foundation
http://dx.doi.org/10.1016/j.jvoice.2013.06.011

in voice, and determining the authenticity of recordings and related tasks. It is what this article is all about.

In turn, Forensic Voice consists of two major elements: the analysis (usually electro-acoustical) of those speech signals that have been transmitted and stored and the examination of communicative acts themselves. The first of these two domains addresses how to properly transmit and store spoken exchanges, the enhancement of speech intelligibility, speech decoding (including accuracy of transcripts), the authentication of recordings and the like. The second area involves recognition of speakers from their voices, identification of the health, emotional or psychological states of the talker, and the analysis of speech for evidence of deception. Naturally, work in these areas interface with a number of other specialties. Included are Forensic Linguistics (especially with respect to language and its analysis), Forensic Psychoacoustics (for auditory-perceptual issues) and sometimes even Audio Engineering. Because these all are tangential to the field, they will not be discussed in this brief review.

SIGNAL ANALYSIS

Good recordings can be of great importance to groups such as law enforcement agencies, the courts, and the intelligence services. Indeed, the effectiveness of most of these agencies would be reduced by a magnitude if they suddenly could no longer use recorded spoken intercourse and related events for surveillance, interrogation, and/or other operations. Indeed, it is quite possible that analysis of such information ranks among the more powerful of the tools investigators have at their disposal. Yet the review and analysis of messages, events, and information has become so common, many investigators, attorneys, and agents tend to overlook the myriad of problems associated with it.

Recordings obtained from surveillance and related activities are rarely of studio quality and the utterances found in many of them will have to be enhanced before they can be understood. The primary sources of speech degradation are related to input noise and distortion; however, there also are problems induced by speakers (ie, overlapping talkers, effects of stress, drug/alcohol usage, and so forth). These disturbances often result from inadequate equipment, poor recording techniques, operator error, and poor recording environments. In turn, these problems create challenges—the first of which involves understanding what the recordings contain.

Speech enhancement

The initial task in decoding messages is to enhance its intelligibility so that the speech can be understood. Various techniques are available to assist with this process.[13–16] They take into account that speech can be, and often is, degraded by: (1) reduction of frequency bandwidth, (2) addition of noise (of any type), (3) reduction in energy level, (4) spectral or harmonic distortion, (5) problems resulting from inadequate transmission links or coupling, (6) inadequate pickup transducers (microphones, "bugs," telephones), (7) operator error, and so on. Masked or degraded speech also can result from environmental factors such as "hum," the wind, vehicle movement, fans/blowers, clothing friction, other talkers, music, or "forensic" noise (see Dallasarra et al.[17] as to how work in forensic voice can be complicated). But, whatever the causes, all detrimental events must be identified and compensatory action taken. The remedies here include several types of electronic filtering, noise elimination (computer) programs,[18,19] restructuring the speech signal, and, of course, decoding by voice specialists.

The initial step in the enhancement process is to protect the original recording by making a good quality digital backup. It is not desirable to work on an original recording because repeated playing could result in signal deterioration or other damage. It is not difficult to imagine the problems that would arise if important evidence were thusly destroyed or compromised. Second, the examiner should listen to the recording a number of times before attempting to process it. This procedure permits development of a log plus a good working knowledge of the recording's contents—*and* information about the interference and degradation present. It is often helpful to digitize problem areas and apply software, which can permit "visualizations" of the relationships within the signal (eg, spectra or waveforms). These graphs also can be used to identify relevant speech sounds, words, or phrases.

Filtering[14,20] by use of relevant analog or digital equipment and/or computer software, often is used as an initial step in the decoding process. If there is a substantial amount of noise at the extreme low and/or high frequencies, speech may often be enhanced by band-pass filters (using the frequency range 300–3500 Hz). If spectrum analyses reveal that a relatively narrow band of energy exists at a specific frequency, a notch filter may be used. When identifiable noise exists *within* the frequency range of the speech, comb filters (especially programmable ones) can be used to continuously modify the signal spectrum by selectively attenuating relatively limited frequency bands. Of course, filtering must be carefully applied so as not to remove necessary speech elements along with the unwanted sounds. Finally, these procedures are best carried out with systems that have been expressly developed to compensate for noise problems.

Because binaural listening can be particularly helpful, the equipment used for speech enhancement should be organized to permit stereo listening by the operator (Bronkhorst[3] re: listening in noise). Furthermore, the use of variable-speed recorders can provide an assist. One type involves a manual increase or decrease of recorder speed while it compensates for parallel distortion of the internal speech signal. Other useful techniques include filling in short dropouts with thermal noise, using linear predictive coefficients, deconvolution, bandwidth compression, and so on. All these procedures are designed to produce a more intelligible speech signal.

Speech decoding

The second phase of the process of extracting utterances and messages from difficult recordings involves speech decoding. Once the exact problems with the recording are identified and speech enhancement techniques have been completed,

Download English Version:

https://daneshyari.com/en/article/1101852

Download Persian Version:

https://daneshyari.com/article/1101852

Daneshyari.com