Review Article

Voice - How humans communicate?

Manjul Tiwari, Maneesha Tiwari

Departments of Oral Pathology and Microbiology, School of Dental Sciences, Sharda University, Greater Noida, Uttar Pradesh, India

Address for correspondence:
Dr. Manjul Tiwari, D-97, Anupam Apartments, B/13, Vasundhara Enclave, Delhi – 110 096, India, E-mail: manjultiw@gmail.com

Abstract

Voices are important things for humans. They are the medium through which we do a lot of communicating with the outside world: our ideas, of course, and also our emotions and our personality. The voice is the very emblem of the speaker, indelibly woven into the fabric of speech. In this sense, each of our utterances of spoken language carries not only its own message but also, through accent, tone of voice and habitual voice quality it is at the same time an audible declaration of our membership of particular social regional groups, of our individual physical and psychological identity, and of our momentary mood. Voices are also one of the media through which we (successfully, most of the time) recognize other humans who are important to us—members of our family, media personalities, our friends, and enemies. Although evidence from DNA analysis is potentially vastly more eloquent in its power than evidence from voices, DNA cannot talk. It cannot be recorded planning, carrying out or confessing to a crime. It cannot be so apparently directly incriminating. As will quickly become evident, voices are extremely complex things, and some of the inherent limitations of the forensic-phonetic method are in part a consequence of the interaction between their complexity and the real world in which they are used. It is one of the aims of this article to explain how this comes about. This subject have unsolved questions, but there is no direct way to present the information that is necessary to understand how voices can be related, or not, to their owners.

Key words: Forensic phonetic, phonetic, sound, voice.

INTRODUCTION

The meaning of "voice"

Perhaps the normal response to the question what is a voice or try to say what voice means. The woeful Oxford English Dictionary (OED) gloss of "voice:—Sound formed in larynx etc. and uttered by mouth, …" is one good example of the semantic inadequacy of dictionary definitions. Another shortcoming is that we cannot assume that the meaning is invariant across languages. After all, many languages, unlike English or French, German, Russian, or Japanese, lack a separate non-polysemous word for voice.[1,2] In Modern Standard Chinese, for example, shingyin is polysemous: It can correspond to either sound or voice. In languages such as this the meaning of voice is not separately lexicalized, but is signaled with sounds that, depending on the language, also mean noise, neck, language, and so on. One should not accord too much significance to whether a language lexicalizes a concept like voice or not: The American Indian language Blackfoot, for example, outdoes most other languages in this regard. It is said to have a verb that specifically means to recognize the voice of someone.[1,2] Let us stick with English, therefore.

Access this article online	
Quick Response Code:	Website: www.jnsbm.org
	DOI: 10.4103/0976-9668.95933

One key semantic component of the English word voice, conspicuously absent from the dictionary definition, must surely be the link with an individual. An acceptable paraphrase of the meaning of the word voice might therefore be vocalizations (i.e., sound produced by a vocal tract) when thought of as made by a specific individual and recognizable as such.[1,3]

Within this semiotic approach there are three important things to be discussed:

- The distinction between voice quality and phonetic quality
- Tone of voice
- The model of a voice

It is possible, at a stretch, to characterize the first two as considering a voice primarily from the point of view (point of hearing, really) of the listener, and the last (i.e., the model) from the point of view of the speaker. To a certain extent the voice model presupposes the voice quality/phonetic quality distinction, so this distinction will be covered first, together with tone of voice. The terms linguistic, extralinguistic, and paralinguistic (or paraphonological) are often used to qualify features functioning to signal phonetic quality, voice quality, and tone of voice, respectively.[1]

VOICE QUALITY AND PHONETIC QUALITY

As already mentioned, when we hear someone talking, we are primarily aware of two things: what is being said, and characteristics of the person saying it. The aspects of the voice that correspond most closely to these two types of judgments on the content and source of the voice are termed phonetic quality and voice quality, respectively.[1,4,5]

Phonetic quality

Phonetic quality refers to those aspects of the sound of a voice that signal linguistic—in particular phonological—information.[1,4,5] In the more technical terms, phonetic quality constitutes the fully specified realizations, or allophones, of linguistic units such as vowel and consonant phonemes.[1,6]

For example, the phonetic quality of the aM vowel phoneme in the word cart as said by an Australian English speaker might be described as long low central and unrounded, and transcribed as [aM]. Some of these phonetic features indicate that it is the linguistic unit, or phoneme, aM that is being signaled, and not some other phoneme.[1]

Linguistic information is being conveyed here, because the choice of a different word is being signaled. Other phonetic features of the vowel are simply characteristic of the range of possible allophones for this phoneme in Australian English. For example, in another speaker, the same vowel phoneme/aM/ might have a phonetic quality described as long low and back of central (M).[1,4,7-9]

Phonetic quality is not confined to segmental sounds such as consonants and vowels, but is also predicated of suprasegmental linguistic categories such as intonation, tone, stress, and rhythm. Thus, the stress difference between INsult and inSULT also constitutes an aspect of phonetic quality, as it signals the linguistic difference between a verb and a noun, as do those aspects of sound that make the following two sentences different: when danger threatens, your children call the police versus when danger threatens your children, call the police.[1,10,11] The linguistic difference being signaled in this example is the location of the boundary between the syntactic constituents of the sentence, and it is being signaled by intentional pitch (rising pitch, a so-called boundary tone, on threatens in the first sentence; rising pitch on children in the second). Many phoneticians would also extend the notion of phonetic quality to those features of sounds that characterize different languages or dialects and make one language/dialect different from another.[1,10,11]

Voice quality

Voice quality is what one can hear when the phonetic quality is removed, as for example when someone can be heard speaking behind a door but what they are actually saying is not audible.[1]

Voice quality is usually understood to have two components: an organic component and a setting component.[1,4-6] The organic component refers to aspects of the sound that are determined by the particular speaker's vocal tract anatomy and physiology, such as their vocal tract length or the volume of their nasal cavity, and which they have no control over. A speaker's anatomical endowment typically imposes limits to the range of vocal features; thus a good example of an anatomically determined feature would be the upper and lower limits of a speaker's fundamental frequency range.[1,12,13]

The second component of voice quality, often called the setting or articulatory setting, refers to habitual muscular settings that an individual adopts when they speak. A speaker may habitually speak with slightly rounded lips, for example, or nasalization, or a low pitch range. Because these setting features are deliberately adopted, they differ from the first component in being under a speaker's control.[1,12,13]

The components of an individual's articulatory setting (e.g., lip rounding) are conceived of as deviations from an idealized neutral configuration of the vocal tract. For example, a speaker might speak with the body of their tongue shifted slightly backward and upward from a neutral position, resulting in what is described as uvularized voice (the deviation being in the direction of the uvula). This would mean that all sound segments susceptible of being influenced by the setting would be articulated further back and slightly higher than normal. The initial stop in the word art, for example, might be articulated further back in the mouth toward the uvular place (qh) instead of at the normal velar, or back of velar, place (kh), and the final alveolar stop [t] might be slightly uvularized (i.e., articulated with the tongue body backed and raised at the same time as making contact at the alveolar ridge (C).[1,3,7,8,9]

The vowel in this example, too, might be articulated further back as [M], thus illustrating an important point. This is that part of the nature of a segment—in this case the backness of the vowel—can often be either the result of a quasi-permanent articulatory setting or simply a lawful allophonic realization of the phoneme in question (it was pointed out above that the Australian English phoneme aM could be realized either as [aM] or [M]). Which one it is—allophone or setting—is shown by the adjacent segments: A pronunciation of [kh, Mt] for cart indicates that the backness of the aM is a phonetic feature; a pronunciation of [qh, MC] indicates that the backness is part of a deliberate voice quality setting. This example shows that whether a particular feature is an exponent of phonetic quality or voice quality depends on how long it lasts: the features as exponents of linguistic segments are momentary; features as realizations of settings are quasi-permanent.[1,5,7,14,15]

The difference between voice quality and phonetic quality can also be illustrated from the point of view of their different roles in the perception of phonetic features. It has been pointed out that voice quality provides the necessary background against which the figure of the phonetic quality has to be evaluated.[4,5] For example, a speaker's linguistic pitch—the pitch that signals the difference between a high tone and a low tone in a tone language—can only be evaluated correctly against the background of their overall pitch range. This was actually demonstrated with the Cantonese tone, where it was shown how the linguistic import of a particular fundamental frequency value—what linguistic tone it was signaling—depended on the speaker's fundamental frequency range, that is aspects of their voice quality. It was pointed out that how vocalic correlates to perceived vowel height, but a particular value for a first formant frequency in a vowel, for example, needs to be evaluated against the speaker's range before it can be decided whether it is signaling a high vowel or a low vowel.[1,15]

Sometimes it is possible to be able to hear both the voice quality and phonetic quality aspects of speech. Thus, if one listens to a male and a female speaker of a tone language saying a word with a high falling tone, it is possible to hear that the phonetic pitch (signaling the tone as high falling) is the same.[1,8,9,12,13] It is also possible to pay attention to the voice quality pitch and hear that, despite the phonetic quality identity, the female has a different, higher voice quality pitch than the male. This is not possible with vocalic quality, outside of specialized techniques such as overtone singing, however. It is possible, for example, to hear that female and male are both saying the same vowel with the same phonetic quality, but it is not possible to hear the accompanying voice quality difference in formant frequencies other than as one of sex.[1,16]

Tone of voice

It is conceivable that, although we cannot hear what our hypothetical post-portal speaker is actually saying, we can hear something about how they are saying it. They may sound angry, for example, or whingeing. The sound features that communicate this information constitute what is called tone of voice, and tone of voice is therefore one of the main ways in which we verbally signal temporary emotional states.[1,17]

It will perhaps come as no surprise that tone of voice shares the same dimensions of sound as phonetic quality and voice quality. Because phonetic quality, voice quality, and tone of voice are all realized in the same dimensions, the question arises of how we perceive the differences. It is assumed that difference between phonetic quality, voice quality, and tone of voice features lies primarily in how long the features are maintained. Tone of voice features are maintained for a duration intermediate between the quasi-permanent voice quality and the momentary phonetic quality features: for as long as the particular attitude is being conveyed.[1,18,19]

THE NEED FOR A MODEL

The sections above have pointed out that when we speak, a lot of information is signaled. This information is informative: it makes the receiver aware of something of which they were not previously aware.[15,20] We obviously think first of all about the information in the linguistic message that we intend to convey. But there are many other types of information, some of them intended, some not.[1]

The different types of information in speech are encoded in an extremely complex way. One aspect of this complexity that with different speakers traces, and different speakers acoustic vowel plots, is that the different types of information in speech are not separately and discretely partitioned, or encoded in separate bits of the message. There is not one frequency band, for example, that signals the speaker's health, or one that signals emotion; the phonetic quality is not a frequency-modulated as opposed to an amplitude modulated voice quality. Such things are typically encoded in the same acoustic parameter.[1,19,21]

Unless the details of this encoding are understood, it is not possible to interpret the inevitable variation between forensic samples. Let us take once again the example of average pitch to illustrate this. As already explained, pitch reflects the size of a speaker's vocal cords, but it also encodes linguistic differences like that between statement

and question, differences in emotion, and differences in health. Unless we understand the details of this encoding, it is not possible to interpret the inevitable pitch variation between samples. An observed difference in pitch might reflect one speaker speaking differently on two occasions (with a preponderance of questions on one occasion, and statements on the other), or two different speakers with different-sized cords speaking in the same way.[1,15,21,22]

The principle involved here is this. Two samples from the same speaker taken under comparable (i.e., totally controlled, as in automatic speaker verification) circumstances are likely to be similar and favor correct discrimination as a same-speaker pair. In the same way, two comparable samples from different speakers are also likely to be correctly discriminated as a different-speaker pair. But non-comparability of samples can lead to incorrect discrimination. It can make two samples from different speakers more similar, thus resulting in evaluation as a same-speaker pair, or it can amplify the difference between two samples from the same speaker, thus favoring evaluation as a different-speaker pair.[1,21-23]

This means that in order to understand how these speaker-specific bits of information are encoded in the speech signal, it is necessary to understand what the different types of information in speech are; what the different components of the voice are; and what the relationship is between the information and the components. To answer these questions now turn to:[1]

VOICE AS "CHOICE" AND "CONSTRAINT"

When we speak it is often because we have information to communicate. However, this information has to be processed through two channels: most obviously, the message has to be implemented by a speaker's individual vocal tract. But the message has to be given linguistic form too, and both these channels affect the form of the message. The result of passing information we want to convey through these channels is the voice.[1,18-24]

When we want to communicate something in speech, we have to make choices within our linguistic system. For example, when we want to signal the word "back" as against "bag" we choose the phoneme k instead of G after bac. When we want to signal our assumption that the hearer can identify the thing we are talking about, we select the definite article "the" instead of the indefinite "a" (the book vs. a book).[1]

But these choices have to be processed through our individual vocal tracts to convert them into speech and therefore are constrained by the physical properties of the individual's vocal tract. This leads to Nolan's characterization of the voice as the interaction of constraints and choices in communicating information. A speaker's voice is the interaction of constraints imposed by the physical properties of the vocal tract and choices that a speaker makes in achieving communicative goals through the resources provided by the various components of his or her linguistic system.[15]

This can be regarded as the picture of the components of a voice. It can be seen that the model consists of four main parts, the connections between which are symbolized by fat arrows. Two of these parts are inputs and two are mechanisms. The two inputs are labeled communicative intent and intrinsic indexical factors, and the two mechanisms are labeled linguistic mechanism and vocal mechanism. The communicative intent maps onto the linguistic mechanism, and the intrinsic indexical factors map onto the vocal mechanism. The vocal mechanism accepts two inputs, from the intrinsic indexical factors and the linguistic mechanism. There is also a picture of a speech wave coming from the vocal mechanism. This represents the final physical, acoustic output of the interaction. This output can be thought of as both the thing that a listener—perhaps best thought of as the forensic phonetician—responds to, and the acoustic raw material that is analyzed by the forensic phonetician.[4,18,25-27]

LINGUISTIC STRUCTURE

As conceived in linguistics, language is a complex multilayered code that links sound and meaning by a set of abstract rules and learnt forms. A very simple model for the structure of this code is shown in Figure 1. As can be seen, it has five components: semantics, syntax, morphology, phonology, and phonetics.[28,29]

Semantics has to do with the meanings conveyed in language; syntax with how words are combined into sentences. Morphology is concerned with the structure of words, and phonetics and phonology encompass aspects of speech sounds. In linguistics, all this structure is termed the grammar of a language, and thus grammar has a wider meaning than is normally understood. The voice's linguistic mechanism in Figure 1 can therefore be properly understood to comprise, in addition to the tone of voice, a large part of the speaker's grammar.[1,6,30,31]

With one proviso described below, in addition to indicating the main components of linguistic structure, Figure 1 can be understood as representing the suite of processes involved when a speaker communicates a specific linguistic

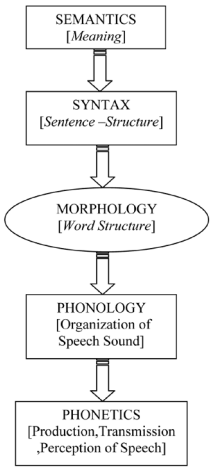

Figure 1: Analytic components of language structure

message verbally to a listener. This is sometimes called "the speech chain," and is symbolized by the bidirectional arrows. Thus, the speaker has a meaning they want to convey, and the meaning is expressed in syntactic, morphological, and, ultimately, acoustic-phonetic form (this is what the downward arrows imply).[1]

It is this acoustic-phonetic form that reaches a listener's ears and that they decode, via their naturally acquired knowledge of the phonetics, morphology, syntax, and semantics of their native language, to reconstruct the meaning of the original message (the implication of the upward pointing arrows). Figure 1 thus represents the linking of speaker's to listener's meaning via sound by showing semantics and phonetics peripherally, joined by the remaining three components of the linguistic code.[1]

These five modules of linguistic structure traditionally constitute the core of any linguistics programmed and are the major categories in terms of which the grammar of a previously undescribed language is described in descriptive linguistics. Because they are also part of the voice, and because they may be referred to in forensic-phonetic reports, it is important to provide a brief characterization of each.[1,2,13,30,31]

SEMANTICS

One of the main differences in the way linguists view the structure of language has to do with the place of meaning: specifically, whether it is primary or not. The view described here will simply assume that it is. That is, as described above, a speaker has meanings they want to communicate, and these are given syntactic, morphological, phonological, and phonetic structure. This is why semantics is placed at the top of the model in Figure 1.

Semantic structure comprises firstly the set of meanings that are available for encoding in language in general and the meanings that have to be encoded in a specific language. For example, all languages allow us to refer to objects and to their location in space.[10,18,23,31] In order to illustrate this, let us assume that someone wish to communicate the location of an object to other person, for example: "The book is over there." In English there are certain semantic aspects that we do not have to encode. We do not have to refer to the book's location being uphill or downhill from the speaker nor to whether the location is near you, or away from you; nor whether the object is visible to me, or you; nor to the source of my knowledge about the book's location, and my consequent belief in its truth. All these are semantic categories that have to be encoded in some languages.[1,6,30-32]

The second type of meaning is structural semantics or the meaning of grammatical structures. As an example of structural meaning, take the two sentences: The man killed the burglar, and The burglar killed the man. The two sentences clearly mean something different, yet they have the same words, so their semantic difference cannot be a lexical one. The difference in meaning derives from the meaning associated with the sentences' syntactic structure: in this case the difference in structural position between the noun coming in front of the verb and the noun coming after. Preverbal position, at least with this verb in this form, is associated with a semantic role called agent: that is the person who prototypically does the action indicated by the verb.[1,30,31]

Postverbal position encodes the semantic role patient. This is prototypically the person who is affected by the action of the verb (in these two sentences the degree of affectedness is extreme, with the patient undergoing a considerable, indeed irreversible, and change of state).[1,30,32]

The third kind of linguistic meaning, pragmatic meaning, has to do with the effect of extralinguistic context on how an utterance is understood. An example of this is the understanding of the meaning "Please give me a bite" from the observation Mm! that looks yummy, the form of which utterance linguistically contains no actual request to carry out the action. As an additional example, pragmatics has to be able to explain how the sentence That's very clever can be understood in two completely opposite ways, depending on the context.[1,25,30-32]

Although semantics is clearly part of linguistic structure, meaning in the voice model is probably best thought of as a part of one of the inputs to the system (the part labeled communicative intent, which is the meaning that the speaker intends to convey), rather than as a part of linguistic structure.

SYNTAX

Syntax functions as a framework on which to hang the structural and pragmatic meanings. Obviously, linguistic meanings have to be conveyed in sequences of words. However, words are not simply strung together linearly, like beads on a string, convey a meaning. They are hierarchically combined into longer units such as phrases, clauses, and sentences, and it is this hierarchical structure that syntax describes.[1,3,15,29]

Syntactic structure is described in terms of constituents, which are words that behave syntactically as a single group. Thus, in the sentence "The exceedingly ferocious dog bit the man," the four words the exceedingly ferocious dog form one constituent (called a noun phrase) for three main reasons. First, the group of words has a particular internal structure, expressed in terms of word class, typical of noun phrase constituents. It consists of an article (the), an adverb (exceedingly), an adjective (ferocious), and a noun. Second, the group can be substituted by a smaller item, for example, the pronoun it, and still yield a grammatical sentence (It bit the man). Third, their constituent status is shown by the fact that they can be moved as a group to form, for example, the related passive sentence: The man was bitten by the exceedingly ferocious dog.[1,11,15,18]

The hierarchical combination of syntactic constituents such as noun phrases into higher order constituents is shown by the fact that the noun phrase the exceedingly ferocious dog forms part of the prepositional phrase by the exceedingly ferocious dog. The resulting structure can be expressed by a syntactic rule such as "prepositional phrase = preposition plus noun phrase." This is one example of what is meant by the structure of the linguistic code being rule-governed.[1]

MORPHOLOGY

The smallest meaningful unit in a language is called a morpheme, and words may consist of one or more morphemes. The single word books, for example, consists of a morpheme meaning "book" and a morpheme meaning "plural." The average number of morphemes per word is one of the main ways in which languages differ. Vietnamese has on average very few morphemes per word; English has on average somewhat more.[1,3,31]

The different types of morphemes and the ways they combine to form words are the subject of morphology. The reader might like to consider how many morphemes are present in the word oversimplification. It consists of four morphemes: A basic adjectival root morpheme simple; a suffix that functions to change an adjective (simple) into a verb (simplify); a prefix over- that attaches onto a verb or adjective (cf. oversub- scribe, overzealous); and a suffixation that functions to change a verb (oversimplify) into an abstract noun. It is not clear whether the -c- in oversimplification is a part of the morphemeify or the morphemeation.[1,23,34]

PHONOLOGY

Phonology deals with the functional organization of speech sounds. One aspect of phonology central to forensic phonetics, namely phonemics recalled that phonemics describes what the distinctive sounds, or phonemes, of a language are, what the structure of words is in terms of phonemes, and how the phonemes are realized, as allophones.[1]

It is one of the interesting structural features of human language that its meaningful units (i.e., morphemes) are not signaled by invariant sounds. An example of this is the English "plural" morpheme called morphophonemics, where it was shown how the plural morpheme is sometimes realized as an [s], as in cats; sometimes as a [z] as in dogs; and sometimes as [mz] as in horses. Which of these forms (or allomorphs) is chosen is predictable and depends on the last sound in the noun. This is another example of the predominantly rule-governed nature of the linguistic code. The area of linguistic structure that is concerned with relationships between the morphemes (meaningful units) and their allomorphs (realizations in sound) is called morphophonemics. It is usually considered as another aspect of phonology.[1,13-15]

PHONETICS

Phonetics deals with the actual production, acoustic realization, and perception of speech sounds.

COMMUNICATIVE INTENT

We now turn to the main input to the system, namely all the information that a speaker intends to convey. This is termed communicative intent.[15] For the speaker to intend something, it has to be the result of a choice (when the characterization above of voice as choice vs. constraint). What sorts of things can and do speakers deliberately encode in their voices?[1]

The first that springs to mind is the linguistic message itself: a proposition (an utterance with a truth value) perhaps,

or a question, or a command. However, speakers also deliberately express emotion; convey social information; express self-image; and regulate conversation with their interlocutor(s), and these also constitute different components of communicative intent. The communicative intent box thus contains five smaller boxes, which refer to these five possible different types of information. These different types will now be described.[1]

COGNITIVE INTENT

Probably the first type of information that one thinks of in speech is its "basic meaning." This is called cognitive information and refers to meaning, differences in which are conveyed by a particular choice and arrangement of words.[1,15]

Because we are dealing with linguistic meaning, changes in cognitive content will have consequences for all the components of linguistic structure. A change in the cognitive meaning of an utterance will be represented in its linguistic semantic structure and result in a change in the selection of a word and/or syntax, and this in turn will cause utterances to be phonologically and phonetically non-equivalent.[1,3,8,15]

AFFECTIVE INTENT

We can choose to signal an emotional state when we speak. Affective intent refers to the attitudes and feelings a speaker wishes to convey in the short term.[15] If we try saying the same sentence in a friendly, then angry tone of voice and note what changes occur. One change will almost certainly be that the angry utterance will be louder, and perhaps the overall pitch will be higher. So another way in which speech samples can differ is in affective intent.[1]

How different emotions are actually signaled in speech is very complicated. More commonly, perhaps, different emotions are signaled linguistically in sound. This occurs primarily by the control of intonational pitch.[1]

We can also signal differences in emotion non-discretely, by for example altering our pitch range. Yes, said with a pitch falling from high in the speaker's pitch range to low signals more enthusiasm than a yes said on a narrower pitch range, with a pitch falling from the middle of speaker's pitch range to low. In these cases, there is a more direct relationship between the actual realization and the degree of emotion signaled, with the degree of involvement reflected in the size of the pitch fall, or the width of the range. Emotion is also commonly signaled in sound by phonation type—the way our vocal cords vibrate.[1,4,5,31,36]

SOCIAL INTENT

Speakers are primates. They interact socially in complex ways. Part of this social interaction is played out in language and is responsible for both between-speaker and within-speaker variations.

It is often assumed that the primary function of language is to convey cognitive information. However, a very important function of language is to signal aspects of individual identity, in particular our membership of a particular group within a language community. This group can be socioeconomically defined. The idea here, then, is that speakers typically choose to signal their membership of social, ethnic, or regional groups by manipulating aspects of linguistic structure. That is part of what is meant by the social intent sub-part of communicative intent.[15]

REGULATORY INTENT

We all talk to ourselves (or the computer, or dog) from time to time, but most[14] involves verbal interaction, usually conversation, with other humans. Conversation is not haphazard: It is controlled and structured, and the conventions underlying conversational interaction in a particular culture are part of the linguistic competence of all speakers who participate in that culture. In traditional Australian aboriginal societies, for example, in contrast to Anglo-Australian culture, it is not normal to elicit information by direct questions. (The obvious implications of this for aboriginal witnesses in court have often been pointed out.) The sub-discipline of linguistics that investigates how speakers manage conversations is called conversation analysis.[1,29,35]

Regulatory intent has to do with the conventional things you deliberately do to participate in a conversation in your culture.[15]

SELF-PRESENTATIONAL INTENT

Richard Oakapple sings in the Gilbert and Sullivan opera Ruddigore: "If you wish in the world to advance/ Your merits you're bound to enhance/You must stir it and stump it and blow your own trumpet/Or trust me you haven't a chance." This reminds us that speakers can deliberately use their voice to project an image to others.[15] This starts early. It is known that little boys and girls, although they are too young to show differences in vocal tract dimensions associated with peripubertal sexual dimorphism, nevertheless choose to exploit the plasticity of their vocal tract in order to sound like (grown-up) males

and females. Little boys have been shown to use lower F0 values and little girls higher.[1]

By their voice, speakers can project themselves as, for example, feminine, confident, extrovert, macho, diffident, and shy. To the extent that this self-image changes with the context, and one might very well encounter such a change between the way any suspect speaks with his mates and the way he speaks when being interviewed by the police, we will find within-speaker variation.[1,2,15,22]

HEALTH AND THE VOCAL TRACT

Any changes in health that affect the size or shape or organic state of the vocal tract, or its motor control, will alter its acoustic output, thus contributing to within-speaker variation.

A speaker's state of health is thus also imprinted on their acoustic output. These intrinsic health-related changes can range from temporary (e.g., a head cold), to periodic (effects of menstrual cycle), to chronic (vocal fold polyp), to permanent (effects of surgery, congenital stutter) and can have the usual consequences of making two different speakers more similar, or the same speaker more different in certain parameters.[1]

An instance of the common health factors that affect the acoustic output is a temporary head cold, which might cause inflammation and swelling in the nasal cavities or sinuses, thus altering their volume and compliance and resonance characteristics. Inflammation and swelling associated with laryngitis might make it painful to stretch the cords too much, and this will temporarily alter a speaker's fundamental frequency values, restricting their range of vibratory values, and making it uncomfortable to reach target values.[1]

Accommodatory changes in tongue body movement associated with different dentures will also affect the resonance pattern of vowels which can result in all sorts of errors in the execution of the complex articulatory plan for the correct realization of linguistic sounds.[1, 8,32-37]

The tongue might not achieve closure for a [d], for example, and some kind of [z]-like fricative might result, or there may be local changes in the rate and continuity of speech. Other factors that interfere with normal motor control and feedback are stress and fatigue.

CONCLUSION

This article has described what a voice is from a semiotic perspective, that is, in terms of the information it conveys. It was motivated by the necessity to understand what can underlie variation in a single speaker's vocalizations in order to correctly evaluate differences between forensic voice samples. It has shown that variations in a speaker's output are a function of two things: their communicative intent (itself a combination of what they want to convey and the situation in which they are speaking) and the dimensions and condition of their individual vocal tract (which impose limits, but not absolute values, to the ranges of phonetic features their language makes use of).

The point has been made elsewhere, and it is worth repeating here, that if the internal composition of a voice appears complex, that is because it is. The voice is complex because there are many things that humans choose to communicate; because the linguistic mechanism used to encode these things is immensely complex; because the mapping between the linguistic mechanism and the communicative intent is complex; because the vocal tract used to implement the complex message involves an enormous number of degrees of freedom; and finally because individual vocal tracts differ in complex ways. All these complexities must be understood if we are to be able to accurately estimate whether differences between forensic samples are between-speaker or within-speaker.

ACKNOWLEDGMENT

I want to give my sincere gratitude and acknowledgement to Philip Rose in preparing this article.

REFERENCES

1. Rose P. Forensic Speaker Identi?cation. 2002 Taylor & Francis, 1st Ed.pp 67-298.
2. Laver J MD. The nature of phonetics'. JIPA 2000; 30/1, 2: 31-6.
3. Baldwin J. 'Phonetics and speaker identi?cation'. Medicine, Science and the Law 1979;19:231-2.
4. Rose P. 'Differences and distinguishability in the acoustic characteristics of Hello in voices of similar-sounding speakers'. Australian Review of Applied Linguistics 1999a; 21/2: 1-42.
5. Laver JMD. 'The semiotic nature of phonetic data' in Laver (1991a), Ch. 10: 162- 70.
6. Laver JMD. 'Voice quality and indexical information' in Laver (1991a), Ch. 9: 147-61.
7. McGehee F. The reliability of the identi?cation of the human voice'. Journal of General Psychology 1937; 17: 249-71.
8. Kreiman J, Papçun G. 'Voice discrimination by two listener populations'. UCLA WPP 1985;61: 45-50.
9. Rose P, Clermont F. 'A comparison of two acoustic methods for forensic discrimination'. Acoustics Australia 2001; 29/1: 31-5.
10. Laver JMD. 'The description of voice quality in general phonetic theory' in Laver (1991a), Ch. 12: 184-208.
11. Ladefoged P. 'Expectation affects identi?cation by listening'. Language and Speech 1978; 21/4: 373-4.
12. Pruzansky S, Mathews MV. 'Talker-recognition procedure based on

analysis of variance'. JASA 1964; 36: 2041-7.
13. Kohler KJ. 'The future of phonetics'. JIPA 2000; 30/1,2: 1-24.
14. Bricker PD, Pruzansky S. 'Speaker recognition' in N. J. Lass (ed.) 1976: 295-326.
15. Kersta LG. 'Voiceprint identi?cation'. Nature 1962;196: 1253-7.
16. Nolan F. 'Forensic phonetics'. Journal of Linguistics 1991; 27: 483-93.
17. Broeders APA, Rietveld ACM. 'Speaker identi?cation by earwitnesses' in Braun and Köster (eds) 1995: 24-40.
18. Shipp T, Doherty E. Hollien H. 'Some fundamental considerations regarding voice identi?cation'. letter to the editor of JASA 1987; 82: 687-8.
19. Doddington GR. 'Speaker recognition - identifying people by their voices'. Proc.IEEE 1985;73/11: 1651-64.
20. LaRiviere C. 'Contributions of fundamental frequency and formant frequencies to speaker identi?cation'. Phonetica 1975; 31: 185-97.
21. Wolf JJ. 'Effcient acoustic parameters for speaker recognition'. JASA 1972; 51: 2044-56.
22. Broeders APA. 'The role of automatic speaker recognition techniques in forensic investigations'. Proc. Intl. Congress Phonetic Sciences 1995; 3: 154-61.
23. Bower B. 'Faces of perception'. Science News 2001;160/1:10-12.
24. Lisker L, Abramson AS. 'A cross-language study of voicing in initial stops: acoustical measurements'. Word 1964; 20: 384-422.
25. Noll AM. 'Short-time spectrum and cepstrum techniques for voiced pitch detection'.JASA 1964; 36: 296-302.
26. Deffenbacher KA. 'Relevance of voice identi?cation research to criteria for evaluating reliability of an identi?cation'. Journal of Psychology 1989;123: 109-19.
27. Foster KR, Bernstein DE, Huber PW. 'Science and the toxic tort'. Science 1993; 261:1509-614.
28. Atal BS. 'Effectiveness of linear predication characteristics of the speech wave for automatic speaker identification and verification'. JASA 1974; 55: 1304-12.
29. Tosi O, Oyer HJ, Lashbrook W, Pedney C, Nichol J, Nash W. 'Experiment on voice identi?cation'. JASA 1972; 51: 2030-43.
30. Durie M, Hajek J. 'A revised standard phonemic Orthography for Australian English vowels'. AJL 1994;14: 93-107.
31. Goggin JP, Thompson CP, Strube G, Simental LR. 'The role of language familiarity in voice identi?cation'. Memory and Cognition 1991;19: 448-58.
32. Wells JC. 'British English pronunciation preferences'. JIPA 1999;29/1: 33-50.
33. Kumar A, Rose P. 'Lexical evidence for early contact between Indonesian languages and Japanese'. Oceanic Linguistics 2000; 39/2: 219-55.
34. Cox F. 'The Bernard data revisited', AJL 1998;18: 29-55.
35. Papçun G, Kreiman J, Davis A. 'Long-term memory for unfamiliar voices'. JASA 1989; 85: 913-25.
36. Elliott JR. 'Auditory and F-pattern variation in Australian Okay: a forensic investigation'.Acoustics Australia 2001; 29/1: 37-41.
37. Goldstein AG, Knight P, Bailis K, Connover J. 'Recognition memory for accented and unaccented voices'. Bulletin of the Psychonomic Society 1981;17: 217-20.29.

How to cite this article: Tiwari M, Tiwari M. Voice - How humans communicate?. J Nat Sc Biol Med 2012;3:3-11.

Source of Support: Nil. **Conflict of Interest:** None declared.