Chapter 2
Psychology of Voice

Abstract The sound of every individual's voice is unique due to the difference in the size and shape of vocal cords. The vocal folds loosen and tighten resulting in a change in pitch, volume, timbre, or tone of the sound produced. Analyzing speech from a physiological perspective, this chapter explores the pitch component of voice and how influential it can be. Interestingly, information regarding prosody, emotions, gender and age is affected by pitch and pitch can help in unconsciously divulging the feelings, moods and emotions. The chapter also enlightens vocal behaviour as a powerful index of emotional and personality markers which are paramount in the extraction of meaningful information from acoustic signals and contribute to a better understanding of the psychology of voice and performance capabilities.

Keywords Physiology · Phonemes · Prosody · Pitch range · Pitch · Emotional markers · Personality markers · Benevolence · Extroversion · Neuroticism

Speech is an information-rich signal created at the vocal cords after travelling through the vocal tract and produced at speaker's mouth. It exploits frequency-modulated, amplitude-modulated and time-modulated carriers to convey information about words, identity, accent, expression, style of speech, emotion and the state of the speaker. It is the most natural form of human communication and is related to human physiological capability and sequence of sound and acoustics known as phonemes. It is essentially a non-stationary signal, but can be divided into sound segments which have some common acoustic properties for a short time interval. The information conveyed by speech is composed of multilayered temporal and spectral variation that includes prosody, gender, age, identity, emotional state etc.

To understand speech as a means of communication, to analyze speech for automatic recognition and extraction of information and to discover some physiological characteristics of the speaker, it is necessary to study how to model speech and its correlates and various aspects of speech processing. Speech coding, synthesis, recognition, understanding, speaker verification and language translation are some of the many speech applications that use fundamentals of linguistics,

S. Johar, *Emotion, Affect and Personality in Speech*,
SpringerBriefs in Speech Technology, DOI 10.1007/978-3-319-28047-9_2

acoustics, pragmatics, speech perception, representation and various speech measures and properties thus making speech processing an extensive theoretical and experimental area of research.

The noise-like air from the lungs is temporally and spectrally shaped by the frequency of the openings and closings of the glottal folds and forms the source signal of the speech. As a result, broadly two types of sounds exist: voiced which are periodic and generated by the vocal cords and unvoiced which are aperiodic and noisy in nature. Due to a steady supply of pressurized air, the vocal cords open and close in a quasi-periodic fashion giving rise to voiced sounds like an 'e'. In case of unvoiced sounds, air passes through some obstacle in the mouth and this obstacle leads to a non-uniform, non-periodic pulse of air.

2.1 Pitch as a Major Auditory Attribute

The periodicity of the glottal pulse and the time-variations of glottal pulse period convey the intent, expressional content, intonation and stress in the speech signals [1]. The time duration of one glottal cycle is defined as the pitch period and its reciprocal is the pitch or fundamental frequency. Pitch is determined by the length, tension, mass of the vocal cords and the sub-glottal pressure. It carries information regarding the prosody or rhythm, emotion, speaking style and accent among many others. The following information is contained in the pitch signal:

(a) Gender classification aims to predict the gender of the speaker by analyzing different parameters of the speech signal. It is mainly conveyed by the pitch value and in part by the vocal tract characteristics. The average pitch for males is about 110 Hz while for females it is about 200 Hz [2].
(b) Emotional states are correlated with particular physiological states, which in turn make predictable effects on speech features, especially on pitch, timing and voice quality. Speech emotion recognition is particularly useful for applications which require natural man-machine interaction and when a person is in a state of anger, joy or fear, the speech is fast, loud and with strong high frequency energy. When someone is sad or bored, slow, low pitched speech with weak high frequency energy is produced. Pitch variation is often correlated with loudness variation. Happiness, distress and extreme fear in voice are also signalled by fluctuations of pitch.
(c) Accents convey information about the status of individual entities in the discourse to indicate their relative salience. It is also largely conveyed by changes in the pitch and rhythm of speech. In addition, a certain type of pitch movement may signal an intonational meaning.
(d) Prosody is a parallel channel of communication for carrying information that cannot be deduced from lexical channel. All aspects of prosody are transmitted by muscle motions and time-variations of pitch have a smooth relationship with these muscle tensions. It gives clues to many channels of linguistic and paralinguistic information and can indicate syntax and people's attitudes and

feelings. Even hand gestures, eyebrow and face motions, can be considered prosody because they carry information that modifies and can even reverse the meaning of the lexical channel.

(e) Age and state of health of a speaker is also related to pitch. The biological fact that the ratio of eye diameter to head diameter varies markedly with age develops connections between the sound shape, meanings or communicative intentions, emotions and affect of the speaker. As a result, a visual estimation of the ratio eye diameter/head diameter is a rough indicator of age and size of speaker.

2.2 Speech Markers

The knowledge of perception of sound and extraction of meaningful data from acoustic signals is paramount to understand the relevance and evolution of audio signal processing. This understanding helps to analyze what comprises the pitch, timbre etc. and what makes some sounds especially natural or artificial. The shortcomings and pitfalls encountered during sound processing can also be studied with this knowledge and thus, suggest various extensions that can be made in storing, producing and modifying speech signals.

Voice has long been considered a measure of emotion and a reflector of personality due to its mature potential to tap individual differences in emotional states and personality dispositions [3]. The understanding of the complex interplay of personality, emotional dynamics and voice production has progressed to a level that many technological advances today support the voice-psychology association. The role of psychological processes among voice- disordered groups has also been long debated and remains a controversial topic of argument.

Speech carries a lot of information over and above the content in the language. The concept of speech markers has been incorporated into the domain of sociolinguistics since 1970s. Most individuals do not have a voluntary control over their personality (age, sex, social class etc.) they present to others. These speech markers are often accompanied by non-linguistic cues permitting interlocutors to communicate emotions, attitudes and intentions about their own as well as other's social states.

2.2.1 Emotional Markers in Speech

Emotions can be expressed in voice at the physiological, the articulatory or the acoustic level. It is intimately connected with cognition and many physiological indices change during emotion arousal [4]. There exist a large number of paralinguistic markers embedded in the acoustic, linguistic and non-verbal content of speech that are intertwined with prosody and semantics and are effective in

distinguishing a large range of emotions over a range of human voices and context, adding naturalness to synthesized speech and thereby facilitating effective emotional speech processing. Since emotion analysis varies with culture, language and even population, it is essentially a multi-faceted approach and improvement in speech emotion recognition performance has been achieved by combining gestural information along with acoustic correlates. Anger, fear, sadness, joy, neutral and surprise are some of the common emotions identified by current speech dialogue and processing systems.

Most of the current methods for measurement and analysis of these cues are intrusive and require specialized equipment and expertise to make explicit and detailed predictions regarding the states conveyed in emotional speech. Studies on emotion may focus on the expression of the emotion by the speaker, the acoustic cues that convey the intended emotion, the perception of these cues and the inference about the expressed emotion. Several studies have explored affect inferences from voice cues in listening tests, where the participants are required to judge the emotions expressed in speech samples using various response formats like forced choice and quantitative ratings. According to Scherer [5], various content-masking procedures that disrupt or degrade individual voice cues can be used to study which voice cues are used by listeners to infer specific emotions.

The existence of various voice profiles for different emotions and the complex nature of voice production process make this task quite challenging to successfully achieve the desired purpose. Inconsistent data regarding voice cues to specific emotions, individual differences among speakers, weak emotional effects and interplay of spontaneous and strategic expressions are some sources of variability that pose practical problems to deduce emotion portrayals. As a result, efforts are being directed towards cross-cultural studies, implementation of multi-modal approaches in emotion expression and intensive research collaboration from psychology, acoustics, engineering and computer science to facilitate better understanding of how emotions are revealed by various aspects of the voice.

2.2.2 Personality Markers in Speech

As it has been mentioned before, the scope of voice-based human machine interaction expands beyond directed dialogue and simple command and control type interfaces. Future machines will need to be able to interpret a specific context, which is determined by many factors including the quality of voice, and produce the respective output. An analysis of the semantic nature of personality traits and interpersonal and intra-personal behaviour dispositions reveals the underlying dimensions regarded as essential determinants of social interactive behaviour.

Controversy that surrounds the concept of personality has forced social and behavioural scientists to debate the nature of personality and its impact on behaviour. Since listeners rely heavily on speech style to attribute personality to the speaker, the possibility of accurate personality inferences from speech remains

questionable [6]. In speech based communication vocal manifestations can be modeled to establish a psychological categorization of personality traits. These manifestations are regarded as speech markers of personality that serve as the basis for personality attribution of the listener corresponding to a specific personality disposition of the speaker [7]. These speech cues reflect the individual differences in cognitive processes of the individuals and the relative dominance of certain emotional and motivational states.

The various prosodic characteristics like pitch level, tempo, speech rate and loudness can be modeled in a number of different ways to convey speaker's affective state and attitude [8, 9]. Speech researchers have demonstrated that emotional states differ in their paralinguistic expression and observers use these vocal cues to judge the personality traits and affective states of the speaker. For example, a speaker's age can be judged by voice alone [10]. Voice quality, pitch and pitch range are the important dimensions on which listeners could base their judgments about speaker age [11]. The pitch measurement varies substantially from childhood to adulthood and is also different for men and women. On the other hand, pitch range appears to remain fairly constant during childhood and increases from adolescence to adulthood. Pitch range is also an important indicator of sex and female range is considerably wider than for men. It may also be stated that the speaker characteristics are relatively permanent as they are closely related to speaker's physiology and anatomy. According to Scherer [7], males tend to have higher pitch levels compared to females and this can be attributed to high degree of arousal in males. Active emotions like anger and happiness are associated with fast tempo, and high pitch, whereas low energy state of sadness attributes to slow tempo lower speech rate and mean pitch. Similarly, major personality dimensions of benevolence and competence are also largely related to pitch and speech rate. Lower pitch and faster speech rate are associated with more credibility and hence, more benevolence [8, 12, 13]. However, deception is strongly related to fundamental frequency of voice and an increased frequency signals false utterances and judges the individual as less truthful. This can also be supported by the fact that stressful situations tend to raise the voice's fundamental frequency. From time to time, correlations of the personality dimensions of introversion-extroversion and emotional stability have attracted various researchers and numerous studies have investigated the prosodic parameters pitch range, pitch level, intensity and tempo to model these dimensions in synthetic as well as natural speech [14–16]. Extroverts are more sociable and interactive and introverts are rather conservative, quiet and shy. On the other hand, emotional stability or neuroticism is an internal state of mind rather than interpersonal reaction. Individuals with high neuroticism are easily overwhelmed by feelings and are said to be less confident and unstable as compared to low neurotics who are more calm and controlled [17]. In the past, it has been found that both these personality dimensions significantly influence an individual's behaviour in a variety of contexts and therefore, there is considerable interest in these traits and their manifestations in behaviour [18–20]. Oberlander and Gill [21] performed Parts of Speech analysis on these two groups and predicted that the neuroticism dimension was more closely related to implicitness

and high neurotics used pronouns and verbs more pervasively. Also, high extroverts used more conjunctions overall and low extroverts preferred more nouns and adjectives.

In order to model personality traits for speech synthesis using different speakers and to identify one or more several defined personalities in dynamic situations, future work will be bound to the availability of data and large databases in order to avoid any influence of the bias of the listener's perspective. Speech is a highly complex interaction of communicative as well as informative characteristics that convey information about the speaker's identity, his emotional state and the situational context. In addition to pitch and rate, additive models that involve other vocal factors must be designed to understand how semantic information is conveyed by the paralinguistic parts of speech. Future personalized speech synthesis systems would require an understanding of how personality is encoded in spoken communication along with refined methods to analyze speech. Moreover, capturing emotional states along with the personalities would facilitate a more holistic system of estimating behaviour from speech. Such parametrical synthesis of speech can be used for diverse commercial applications to indicate personality impressions that individuals leave on one another and highlight the existence and psychological significance of personality as an important correlate in social interaction.

References

1. Kashem (2004) Speech processing. http://duet.ac.bd/drkashemweb/Dr.Kashem%20Wev/dr%20kasem-dsp-ps/Chapter13-Speech%20Processing.pdf. Accessed 17 Jun 2015
2. Kawahara H, Matsui H (2003) Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. In: Proceedings of IEEE international conference on acoustics, speech and signal processing, vol I, pp 256–259
3. Aronson AE (1990) Clinical voice disorders: an interdisciplinary approach, 3rd edn. Thieme, New York
4. Lindsay PH, Norman DA (1972) Human information processing. Academic Press, New York and London
5. Scherer KR (2003) Vocal communication of emotion: A review of research paradigms. Speech Commun 40:227–256
6. Giles H, Powesland PF (1975) Speech style and social evaluation. Academic Press, New York
7. Scherer KJ (1979) Personality markers in speech. In: Scherer KR, Giles H (eds) Social markers in speech. Cambridge University Press, Cambridge p, pp 147–209
8. Apple W, Krauss RM (1979) Effects of pitch and speech rate on personal attributions. J Appl Soc Psychol 37:715–727
9. Trouvain J, Barry WJ (2000) The prosody of excitement in horse race commentaries. Proceedings of ISCA—workshop on "speech and emotion". Newcastle, Northern Ireland, pp 86–91
10. Allport GW, Cantril H (1934) Judging personality from voice. J Soc Psychol 5:37–55
11. Helfrich H (1979) Age markers in speech. In: Scherer KR, Giles H (eds) Social markers in speech. Cambridge University Press, Cambridge p, pp 63–108
12. Smith B, Brown B, Strong W, Rencher A (1975) Effects of speech rate on personality perception. Lang Speech 18:145–152

13. Brown BL, Strong WJ, Rencher AC (1974) Fifty-four voices from two: the effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. J Acoust Soc Am 55:313–318
14. Scherer KR, Scherer U (1981) Speech behaviour and personality. Speech evaluation in psychiatry. Grune & Stratton, New York, pp 115–135
15. Nass C, Lee KM (2001) Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. J Exp Psychol Appl 7(3):171–181
16. Mairesse F, Walker MA, Mehl MR, Moore RK (2007) Using linguistic cues for the automatic recognition of personality in conversation and text. J Artif Intell Res (JAIR) 30:457–500
17. Eysenck H, Eysenck SBG (1991) The Eysenck personality questionnaire-revised. Hodder and Stoughton, Sevenoaks
18. Isbister K, Nass C (2000) Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. Int J Hum Comput Stud 53:251–267
19. Furnham A (1990) Language and personality. In: Giles H, Robinson W (eds) Handbook of language and social psychology. Wiley, Chichester, pp 73–95
20. Dewaele JM, Furnham A (1999) Extraversion: the unloved variable in applied linguistic research. Lang Learn 49:509–544
21. Oberlander J, Gill AJ (2004) Individual differences and implicit language: personality, parts of speech and pervasiveness. In: Proceedings of the 26th annual conference of the cognitive science society. Chicago, IL, USA

http://www.springer.com/978-3-319-28045-5

Emotion, Affect and Personality in Speech
The Bias of Language and Paralanguage
Johar, S.
2016, VII, 52 p. 3 illus., Softcover
ISBN: 978-3-319-28045-5