Journal of Theoretical and Applied Information Technology
31st August 2015. Vol.78. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

USING FUZZY SETS TO MODEL PARALINGUISTIC CONTENT IN SPEECH AS A GENERIC SOLUTION FOR CURRENT PROBLEMS IN SPEECH RECOGNITION AND SPEECH SYNTHESIS

SACHIN LAKRA[1], T. V. PRASAD[2], G. RAMAKRISHNA[3]
[1]Research Scholar, Computer Science & Engineering,
K L University, Vaddeswaram, Guntur, Andhra Pradesh, India.
[2]Dean, Research & Planning, Chirala Engineering College, Chirala, Andhra Pradesh, India.
[3]Professor, Computer Science & Engineering,
K L University, Vaddeswaram, Guntur, Andhra Pradesh, India
sachinlakra@yahoo.co.in, tvprasad2002@yahoo.com, ramakrishna_10@yahoo.com

Abstract

Current problems in speech processing exist due to infinite variations of speech utterances. No two speech utterances are exactly alike, even if they are linguistically the same word. The difference is therefore, due to the paralinguistic content of the speech utterances. This leads to the conceptualization of the paralinguistic content of speech as arising from infinite variation. Infinite variation in paralinguistic content has been modeled using the interval [0, 1], the basis of fuzzy theory. Variability, that is, the ability to vary, has been identified as the property of natural systems, due to which infinite variation is possible. Thus, variability as a concept has been mapped to paralinguistic content. Further, each component of paralinguistic content has been mapped to a group of membership functions of fuzzy sets.

Keywords: Variability, Paralinguistic Content Elimination, Paralinguistic Content Expression, Speech Recognition, Speech Synthesis.

1. INTRODUCTION

Natural systems, other than those which are man-made or artificial, all exhibit infinite non-repeating variations in their dynamics. Man-made systems, however, exhibit repetition in their dynamics. Thus, natural systems follow non-linear behaviour, whereas man-made systems exhibit linear behavior. In linear behavior, since there is no requirement for decision-making, therefore, the man-made system can continue to be stable without the requirement of any intelligent behavior. However, since natural systems, such as living beings, come across events in which decision-making is necessary due to choices, therefore, they exhibit intelligent behavior.

An example of non-linear behavior is that of speech processing, that is, speech recognition and speech synthesis. Current problems in speech processing are persisting due to the infinite variation in spoken utterances, on the one hand, and the need for naturalness in speech synthesis on the other hand. Therefore, considering human beings to be natural systems, as distinct from man-made systems, it is essential to overcome infinite variations in speech recognition or introduce infinite variations in speech synthesis so as to solve the existing problems in speech processing.

With this objective, a theoretical foundation of the solution to infinite variation is presented in this paper, along with applying this foundation to develop a model for solving the problem of variations in speech.

2. VARIABILITY

Variability is the ability to vary. Variability is universally found in nature. No two objects are found to be exactly the same, that is, they may differ in terms of structure, behavior, and spatio-temporal characteristics.

Definition 1: Variability is the property of a variable V to be assigned a value $v \in [\alpha, \beta]$.

For example, a variable x, $x \in$ N can be assigned a single value out of an infinite number of values lying in the set of natural numbers N. Thus, x exhibits the property of variability.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.1 Variability Is Necessary

Variability lends uniqueness to objects, events and interactions. Uniqueness in objects, events and interactions leads to spatio-temporal change. This leads to richness in life and, in fact, change is the universal constant event in natural systems.

2.2 A universe without variability

Consider a world without variability, implying a world without differences, and hence without change. There will be no objects which are different, no locations which are different, no difference in behavior, in fact, no change in instances of time. That is, everything will be static; the world will be devoid of dynamics.

2.3 Variability is Infinite

Variability in nature is infinite. For example, no two human beings have exactly the same face, although, each human face has the same basic pattern. That is, each human face has two eyes, one nose, one mouth, etc. Similarly, the human body has many patterns which are unique to an individual, e.g., fingerprints, retina, ear lobes, etc. Thus, each pattern that can be used as a means of uniquely identifying an individual can be said to exhibit variability.

2.4 An interval-based reasoning for uncountability

Consider two real numbers, at random, forming a real interval [0.1, 2.4]. It is possible to identify another interval within this interval.

0.1					2.4
0.1	0.25			2.23	2.4
0.1	0.25	0.259	1.51	2.23	2.4

This process can continue indefinitely for a real interval, theoretically, till infinity. At infinity, the lower and upper bounds of the interval will become equal, thus forming a point.

Thus, there can be an infinite number of intervals within any given interval, e.g., the interval [0.1, 2.4] as above.

Based on this line of thought, an interval-based proof is presented in Section 2.4.1.

2.4.1 An interval-based alternative proof for Cantor's Uncountability Theorem[1], [2]

Theorem 1: A finite interval $[\alpha, \beta], \alpha, \beta \in \mathbb{R}$ consists of an infinite number of intervals.

Proof:

Let x be a point such that $x \in [\alpha, \beta]$.

Let $\Delta x \to 0$, be a very small value.

Let i be the interval $[x - \Delta x, x + \Delta x]$.

Let there be n intervals, each interval of length $2\Delta x$, in $[\alpha, \beta]$, where $n \in \mathbb{N}$, such that

$$\bigcup_{\mu=1}^{n} i_{\mu} = [\alpha, \beta]$$

As $\lim_{\Delta x \to 0} i = [x - \Delta x, x + \Delta x] = x$,

$\Rightarrow n \to \infty$.

Hence, QED.

Thus, a finite interval consists of an infinite number of intervals. Further, as the size of the interval approaches an infinitesimal size, the length of the interval approaches zero and becomes a point in the limit.

Consider a function f,

$$f: x \to y$$

where, $x \in [0, 1]$, then any one of an infinite number of possible values in the interval $[0, 1]$ can be mapped to y.

Thus, at any event, such as the mapping of a single value from an interval to a variable, an infinite number of choices are theoretically available for being mapped.

Hence, there exist an infinite number of possible variations for a given mapping. That is, there exists infinite variability.

Corollary 1 *:* An infinite number of points exist in the interval [0, 1].

3. CURRENT PROBLEMS IN SPEECH RECOGNITION AND SPEECH SYNTHESIS

Current speech technology is lacking in the features which are categorized under naturalness. From Table 1, it is observable that the fundamental problem in all the existing problems in speech processing is the problem of infinite variability. Human beings utter natural speech due to the fact that they produce non-repeating instances of speech. This is true even for the same word to be spoken by different speakers. Thus, to solve each of the problems[3] listed below, the solution is to apply the concept of variability to model this infinite variation.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. VARIABILITY IMPLIES CONTINUA

A continuum is defined as in [4], [5]. Modelling variability mathematically requires an infinite-valued interval, which is most suitably the interval [0, 1]. The interval [0, 1] is already the basis of fuzzy sets. Specifically, this interval can be used to allow comparisons between degrees to which two given crisp values from a universe of discourse can be classified to lie in the same fuzzy set, but with different degrees of membership.

4.1 Representing Infinite Variability In A Continuum As A Membership Function

A continuum can be represented by a multi-valued logic, such as fuzzy logic, since fuzzy logic uses the interval [0, 1] to represent degrees of membership of a crisp interval in a fuzzy set, using a membership function. For example, a continuum such as the frequency spectrum of white light is possible to be represented as a set of 7 linguistic values, that is, {Violet, Indigo, Blue, Green, Yellow, Orange, Red}. Each linguistic value is a fuzzy set which can be represented by a membership function as shown in Fig. 1.

Therefore, it is possible to represent variability with a membership function. Populations of natural data, such as the heights of 1,000 human beings, comply with a normal distribution, which is represented by a Gaussian bell-shaped curve. No two human beings have precisely the same height. Hence, the Gaussian bell-shaped curve is suitable to model natural variability, e.g., that in natural speech.

Therefore, it is possible to represent variability with a membership function. Populations of natural data, such as the heights of 1,000 human beings, comply with a normal distribution, which is represented by a Gaussian bell-shaped curve. No two human beings have precisely the same height. Hence, the Gaussian bell-shaped curve is suitable to model natural variability, e.g., that in natural speech.

The interval [0, 100], used to model percentages, represents, for example, the degrees of completion of a task.

The interval [0, 100], used to model percentages, represents, for example, the degrees of completion of a task. Similarly, the interval [0, 1] is used to model degrees of membership of any crisp interval in the case of fuzzy theory, as a membership function of a fuzzy set.

By Theorem 1, the interval [0, 1] contains an infinite number of points and can model infinite variations.

Table 1. Current Problems In Speech Processing And The Form Of Variability Inherent In Each Problem.

Stage	Current Problems	Form of Variability
Speech-to-text conversion	Noise	No two noise signals are exactly alike
	Speaker variability including realization, speaking style, gender, speed of speaking	No two speakers utter the same speech component in exactly the same manner
	Emphasis	No two instances of emphasis are exactly same even for the same emotion being expressed
	Accent	No two persons from different regions utter speech in exactly the same way
Text-to-Speech conversion	Accent	No two natural speakers from different regions speak a language in exactly the same way
	Phoneme generation with context-specific pronunciation	No two instances of uttering a sound in two different contexts are exactly same
	Segment duration	No two natural speakers utter a sound in exactly the same duration
	Intonation	No two instances of the tone in which singers sing are exactly same
	Intelligibility	Since no two instances of utterances of sounds by one or more speakers is the same their intelligibility is also not exactly same
	Mimicry of a speaker’s voice	No two instances of performing the mimicry of a speaker by another speaker are perfectly same since no two humans are exactly the same
	Voice “Naturalness”	Naturalness of voice itself implies variation since otherwise speech will be monotonous

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

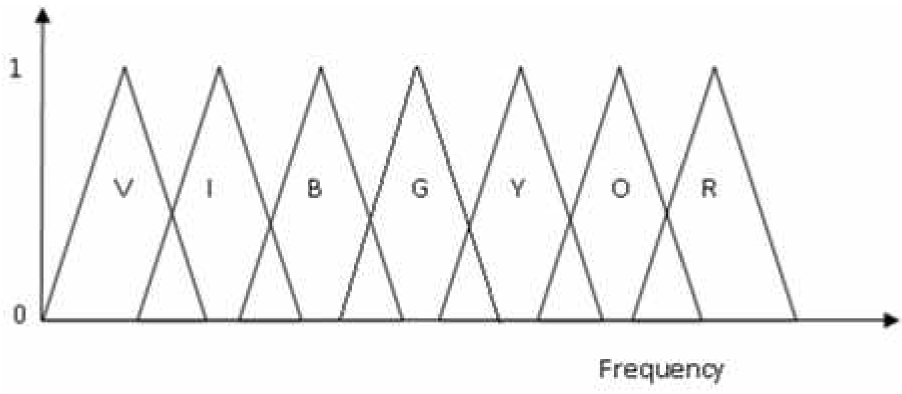

Figure 1. The linguistic variable frequency spectrum of white light represented as a set of linguistic values, that is, {V, I, B, G, Y, O, R}, each represented by a triangular membership function.

5. MODELLING PARALINGUISTIC CONTENT IN SPEECH USING VARIABILITY

5.1 Linguistic vs Paralinguistic Content of Speech

Linguistic content of speech is the textual version of an utterance, e.g., if a sentence spoken by a person is converted to text, then the text is the linguistic content. On the other hand, all the content of speech other than the linguistic content is paralinguistic content, e.g., sounds, accent, gender, emphasis, duration, naturalness, context and semantics.

5.2 Mapping Paralinguistic Content To A Fuzzy Set[6]

Human beings possess the ability to fathom variability by using fuzziness to represent degrees of variability in any paralinguistic parameter. According to fuzzy theory, a continuum can be represented using the interval [0, 1]. Each variation of a paralinguistic parameter can be considered as a value lying in the interval [0, 1]. This value can then be mapped to the membership function of a fuzzy set and classified as a member of the fuzzy set.

$$\mu_A : X \to [0,1] \qquad (3.1)$$

where X is a crisp set representing the values of a paralinguistic parameter, and μ is a membership function mapping x ∈X to the interval [0, 1] and to the fuzzy set A.

6. MODELLING PARALINGUISTIC CONTENT

A spoken utterance has two components to it, one linguistic and the other paralinguistic. The basic element of a spoken utterance is the phoneme. Any one type of paralinguistic component is mapped to the linguistic component as a fuzzy mapping.

For example, consider the phoneme "u" uttered by n different speakers. Then the paralinguistic component of each speaker's utterance is a variation of the phoneme. However, the linguistic component of each variation is the same, that is, the alphabet "u". Thus, there is a degree of membership of each variation in the fuzzy set representing the phoneme. This degree is represented by a fuzzy value in the interval [0,1].

Generalizing, if *P* is a fuzzy set representing one element *p* of the paralinguistic content of a phoneme $\pi \in \Pi$, then the degree of membership of a variation $v \in V$ of π in *P* is given by

$$\mu_P : V \to [0,1] \qquad (3.2)$$

Each phoneme π is mapped to an alphabet $\alpha \in A$ in a language Λ by the function *p*. Here, A is the set of alphabets of a language Λ. Therefore, *P* is also mapped to A. Thus,

$$p : P \to A \qquad (3.3)$$

This is presented in Fig. 2.

6.1 Paralinguistic Content Elimination: Speech Recognition

The problem of speech recognition is, thus, stated as the elimination of one or all components of paralinguistic content from speech to obtain the underlying linguistic content, that is, textual word. The model for paralinguistic content elimination is presented in Fig. 3.

Journal of Theoretical and Applied Information Technology
31st August 2015. Vol.78. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6.2 Paralinguistic Content Expression: Speech Synthesis

The problem of speech synthesis is, thus, stated as the expression of non-repeating variations of one or all components of paralinguistic content to synthesize speech from the given linguistic content, that is, the underlying word. The model for paralinguistic content expression is presented in Fig. 4.

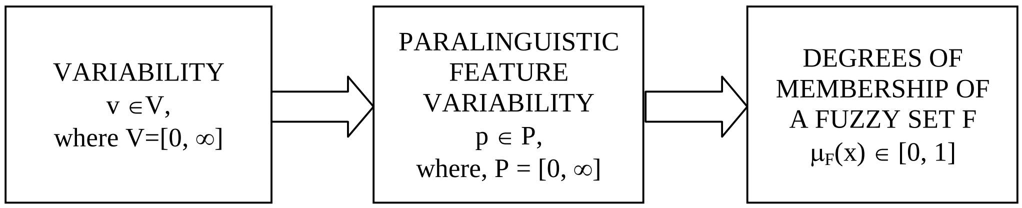

Figure 2. Modelling paralinguistic content using variability as a fuzzy set.

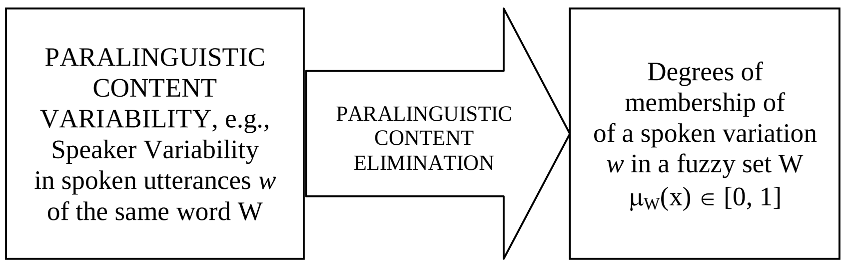

Figure 3. Modelling speech recognition as paralinguistic content elimination

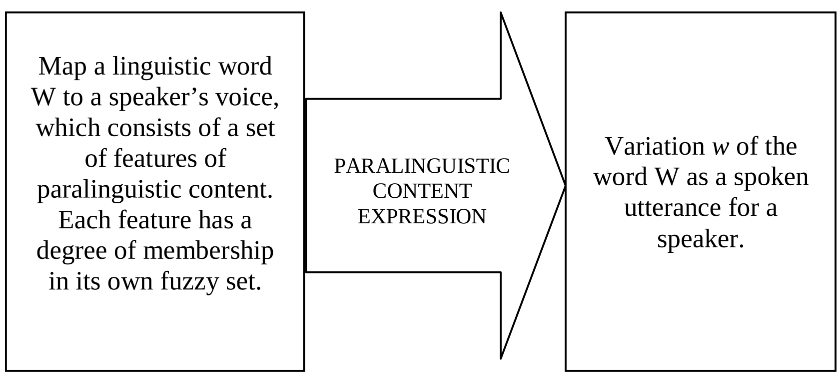

Figure 4. Modelling Speech Synthesis As Paralinguistic Content Expression.

7. CONCLUSIONS

Naturalness will thus be achieved in synthesized speech by using paralinguistic content expression leading to non-repetitive utterances of the same speech signal which is also how human beings express naturalness in speech. Variability elimination and variability expression, as applicable to speech, lead to the operations of paralinguistic content elimination and paralinguistic content expression, for speech processing.

The actual implementation along with results are part of work that is under progress. Using the two operations for paralinguistic content elimination and expression, a model is under development for solving, specifically, the three problems of speaker variability, gender variability and duration variability in speech processing.

REFERENCES:

[1] Mueckenheim, W., "On Cantor's important proofs", *arXiv preprint math/0306200*(2003).

[2] Mueckenheim, W., "On Cantor's important theorems." (2004).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[3] Forsberg, M., "Why is Speech Recognition Difficult?", Chalmers University of Technology, February, 2003.
[4] Stevens, S. S., "On the theory of scales of measurement" (1946).
[5] Feferman, S., "Conceptions of the Continuum", Intellectica (51), 2009, pp. 169-189.
[6] Lakra, S., Sharma, D.K., Prasad, T.V., Atrey, S.H., "Application Of Fuzzy Mathematics To Speech-To-Text Conversion By Elimination Of Paralinguistic Content", Proceedings of National Conference on Soft Computing & Artificial Intelligence 2009, Lingaya's University, Faridabad, Jan 2009.