
Comparative Analysis in Conflict Resolution: Computer Speech Synthesis and Humans Speech Production in View

Okpala Izunna Udebuana

Department of Communication and Translation Studies, National Institute for Nigerian Languages Aba, Akpaa, Nigeria

Email address:

izunna.okpala@yahoo.com

To cite this article:

Okpala Izunna Udebuana. Comparative Analysis in Conflict Resolution: Computer Speech Synthesis and Humans Speech Production in View. *International Journal of European Studies*. Vol. 3, No. 1, 2019, pp. 34-38. doi: 10.11648/j.ijes.20190301.16

Received: March 1, 2019; **Accepted:** April 11, 2019; **Published:** May 10, 2019

Abstract: Conflict is inevitable when it comes to communication between people from diverse background and settings. Computer systems also experience conflicts in form of bugs. Most naturally, before conflict of any sort occurs, be it ideas or perception, there must be some form of communication. Speech is one of the oldest and most natural means of information exchange between human beings. Humans speak and listen to each other in human-human interface in order to resolve certain conflicts, but computers speak to humans in a computer-human interface. The echo that comes out of a given speech might be understood or perceived differently when presented to different people. This paper is based on a comparative approach and focuses on given a run-down of the successes recorded in conflict resolution using human speech production in contrast to computer speech production. The author registers the conflict resolution practices in computers using a try-catch block pseudocode, its effectiveness in conflict resolution, plus the properties it lacks, and then, compares it to that of human functions as regards conflict resolution, in order to find a better approach. The methodology employed in this research is qualitative in nature. The author explores the stages and techniques of applying an artificial Intelligence system that scans through a given speech production and also how the brain processes information before it is finally voiced out.

Keywords: Artificial Intelligence, Information Exchange, Speech Production, Conflict Resolution

1. Introduction

Speech as one of the oldest and most natural means of information exchange between human beings is very important in conflict resolution. Humans speak and listen to each other in human-human interface. For centuries people have tried to develop machines that can understand and produce speech as humans do so naturally. Obviously such an interface yields great benefits. Attempts have been made to develop vocally interactive computers to realize voice/speech synthesis. In this case a computer can synthesize text and give out a speech. More so, Languages are central to exchange of information amongst individuals, groups and the global population. Linguists of all persuasions seem to agree that a language should be viewed as a system; a set of elements, each of which has a capacity of contributing to the workings of the whole [1]. It is through languages that people express their folk tales, myths, proverbs and history [2].

Artificial speech has been a dream of the humankind for centuries. The history of speech synthesis machines began at

least as early as the 17th century [3]. Unfortunately, these early attempts did not leave anything but indirect documentary evidence of their existence. Nevertheless, the early machines were an omen of things to come. Until the late 19th century speech synthesis was based on construction of physical models, which can be considered simple articulated speech synthesizers. The late 19th century gave birth to all sorts of electrical devices and among them was an electromechanical speech synthesizer by Helmholtz. One of the first electrical device analogous to the human speech organs was presented by Stewart (1922), who appears not to have known of Helmholtz's earlier vowel synthesis machine. Stewart's device consisted of an interrupted source and two adjustable RCL-resonator circuits. These can be seen to correspond to the vocal folds (the source and interrupter) and the VT (the resonators).

The modern day computer system is embedded with thousands of chips for speech production. This paper focuses on the comparative analysis between humans and computers on how they handle conflict during speech production – addressing the key factors that initiates conflict, the way

various faculties like the brain and inert feelings influences human speech production and how a programmatic system influences computer speech production.

2. Methodology

The methodology employed in this paper is qualitative and comparative in nature. Here, comparisons are based on the quality and formality of how conflicts are resolved. This research is designed to gain insight and compare the quality of conflict resolution using two instances (human communication and computer communication). The researcher analyses the intangible and inexact concerns that belongs to the social and experiential realm and those of mathematical theories that works in computing technology, given that there are exact data fed into the computer systems, which are not tangible or discrete per se but continuous in nature. This approach checks the kind of intelligence that machines (currently) lack or exhibits, since things like positive associations with a brand, conflict resolution, management trustworthiness, customer satisfaction, competitive advantage and cultural shifts are difficult and arguably impossible to capture with numerical inputs using computer synthesized approach. Without conclusions yet, this paper will explore the two.

3. Theoretical Framework

As a defined field of study, conflict resolution started in the 1950s and 1960s. This was at the height of the Cold War, when the development of nuclear weapons and the conflict between the superpowers seemed to threaten human survival. A group of pioneers from different disciplines saw the value of studying conflict as a general phenomenon, with similar properties whether it occurs in international relations, domestic politics, industrial relations, communities or families or between individuals. They saw the potential of applying approaches that were evolving in industrial relations and community mediation settings to conflicts in general, including civil and international conflicts. A handful of people in North America and Europe began to establish research groups to develop these new ideas. They were not taken very seriously. The international relations profession had its own categories for understanding international conflict and did not welcome the interlopers. Nor was the combination of analysis and practice implicit in the new ideas easy to reconcile with established scholarly institutions or the traditions of practitioners such as diplomats and politicians. Nevertheless, the new ideas attracted interest, and the field began to grow and spread. Scholarly journals in conflict resolution were created. Institutions to study the field were established, and their number rapidly increased. The field developed its own subdivisions, with different groups studying international crises, internal wars, social conflicts and approaches ranging from negotiation and mediation to experimental games. By the 1980s, conflict resolution ideas were increasingly making a difference in real conflicts. In

South Africa, for example, the Centre for Intergroup Studies was applying the approaches that had emerged in the field to the developing confrontation between apartheid and its challengers, with impressive results. In the Middle East, a peace process was getting under way in which negotiators on both sides had gained experience both of each other and of conflict resolution through problem-solving workshops. In Northern Ireland, groups inspired by the new approach had set up community relations initiatives that were not only reaching across community divides but were also becoming an accepted responsibility of local government. In war-torn regions of Africa and South-East Asia, development workers and humanitarian agencies were seeing the need to take account of conflict and conflict resolution as an integral part of their activities.

3.1. Human Speech Production

The figure below (Figure 1) shows the human speech production organs along with an idealized model. The picture on the left shows the main articulators - the tongue, the jaw and the lips - as well as other important parts of the vocal tract (VT). The picture on the right shows the model, which is the basis of almost every acoustical model of the vocal tract. In production of pulmonic sounds breathing muscles act as an energy source while the lungs provide storage of pressurized air. The lungs are separated from the vocal tract by the vocal folds, which are also known as vocal chords. The vocal folds generate a signal, which is then filtered by the vocal tract and finally radiated to the surroundings via the mouth and/or nostrils.

While the above description of the speech production process is fairly accurate for (some stages of) vowel production, it is only a starting point to understanding speech production in general. The neurological process, which leads to articulation movements, is not a part of the described process.

Human speech is produced by vocal organs presented in Figure 1. The main energy source is the lungs with the diaphragm. When speaking, the air flow is forced through the glottis between the vocal cords and the larynx to the three main cavities of the vocal tract, the pharynx and the oral and nasal cavities. From the oral and nasal cavities, the air flow exits through the nose and mouth, respectively. The V-shaped opening between the vocal cords, called the glottis, is the most important sound source in the vocal system. The vocal cords may act in several different ways during speech. The most important function is to modulate the air flow by rapidly opening and closing, causing buzzing sound from which vowels and voiced consonants are produced. The fundamental frequency of vibration depends on the mass and tension and is about 110 Hz, 200 Hz, and 300 Hz with men, women, and children, respectively. With stop consonants the vocal cords may act suddenly from a completely closed position, in which they cut the air flow completely, to totally open position producing a light cough or a glottal stop. On the other hand, with unvoiced consonants, such as /s/ or /f/, they may be completely open. An intermediate position may also occur with for example phonemes like /h/.

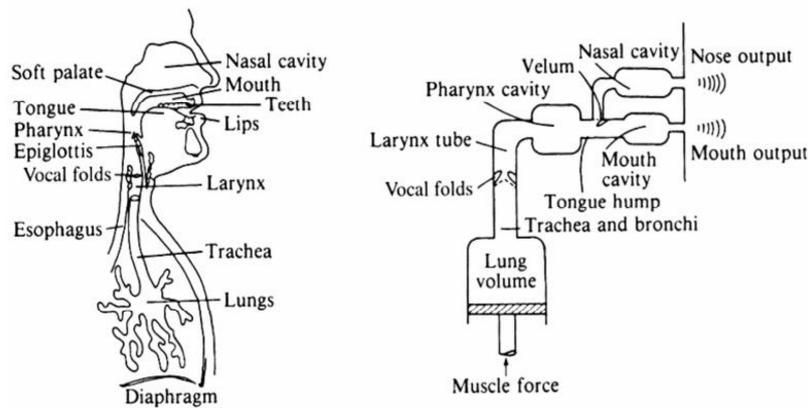


Figure 1. Human speech production components and activity presentation [4].

The pharynx connects the larynx to the oral cavity. It has almost fixed dimensions, but its length may be changed slightly by raising or lowering the larynx at one end and the soft palate at the other end. The soft palate also isolates or connects the route from the nasal cavity to the pharynx. At the bottom of the pharynx are the epiglottis and false vocal cords to prevent food reaching the larynx and to isolate the oesophagus acoustically from the vocal tract. The epiglottis, the false vocal cords and the vocal cords are closed during swallowing and open during normal breathing.

The oral cavity is one of the most important parts of the vocal tract. Its size, shape and acoustics can be varied by the movements of the palate, the tongue, the lips, the cheeks and the teeth. Especially the tongue is very flexible, the tip and the edges can be moved independently and the entire tongue can move forward, backward, up and down. The lips control the size and shape of the mouth opening through which speech sound is radiated. Unlike the oral cavity, the nasal cavity has fixed dimensions and shape. Its length is about 12

cm and volume 60 cm³. The air stream to the nasal cavity is controlled by the soft palate.

3.2. Machine Speech Production

Vocal sounds are inherently governed by the partial differential equations of fluid mechanics, applied in a dynamic case since the lung pressure, glottis tension, and vocal and nasal tracts configuration evolve with time. These are controlled by human cortex, which takes advantage of the power of its parallel structure to extract the essence of the text read: its meaning. Even though, in the current state of the engineering art, building a Text-To-Speech synthesizer on such intricate models is almost scientifically conceivable (intensive research on articulatory synthesis, neural networks, and semantic analysis give evidence of it), it would result anyway in a machine with a very high degree of (possibly avoidable) complexity, which is not always compatible with economic criteria.

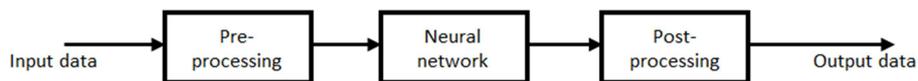


Figure 2. Schematic illustration of input and output of machine speech production.

3.3. Ways Humans Handle Conflict

In the late 1960s Johan Galtung (*Galtung's models of conflict, violence and peace*) proposed an influential model of conflict that encompasses both symmetric and asymmetric conflicts [5]. He suggested that conflict could be viewed as a triangle, with contradiction (C), attitude (A) and behaviour (B) at its vertices. Here the contradiction refers to the underlying conflict situation, which includes the actual or perceived 'incompatibility of goals' between the conflict parties generated by what Chris Mitchell calls a 'mismatch between social values and social structure' (1981). In a symmetric conflict, the contradiction is defined by the parties, their interests and the clash of interests between them. In an asymmetric conflict, it is defined by the parties, their relationship and the conflict of interests inherent in the relationship. Attitude includes the parties' perceptions and

misperceptions of each other and of themselves. These can be positive or negative, but in violent conflicts parties tend to develop demeaning stereotypes of the other, and attitudes are often influenced by emotions such as fear, anger, bitterness and hatred. Attitude covers emotive (feeling), cognitive (belief) and conative (desire, will) elements. Analysts who emphasize these subjective aspects are said to have an expressive view of the sources of conflict (for example: a social conflict exists when two or more parties believe they have incompatible objectives). Behaviour is the third component. It can involve cooperation or coercion, gestures signifying conciliation or hostility. Violent conflict behaviour is characterized by threats, coercion and destructive attacks. Analysts who emphasize objective aspects such as structural relationships, competing material interests or behaviours are said to have an instrumental view of the sources of conflict (for example: there is conflict whenever incompatible actions

occur. An action that is incompatible with another action prevents, obstructs, interferes, injures or in some way makes the latter less likely to be effective).

Galtung argued that all three components (contradiction, attitude and behaviour) has to be present together in a full conflict. A conflict structure without conflictual attitudes or behaviour is a latent (or structural) one. Galtung sees conflict as a dynamic process in which structure, attitudes and behaviour are constantly changing and influencing one another. As the dynamic develops, it becomes a manifest conflict formation, as parties' interests clash or the relationship they are in becomes oppressive. Parties then organize around this structure to pursue their interests. They develop hostile attitudes and conflictual behaviour. And so the conflict formation starts to grow and intensify. As it does so, it may widen, drawing in other parties, deepen and spread, generating secondary conflicts within the main parties or among outsiders who get sucked in. This often considerably complicates the task of addressing the original, core conflict. Eventually, however, resolving the conflict must involve a set of dynamic changes that mean de-escalation of conflict behaviour, a change in attitudes, and a transformation of the relationships or clashing interests that are at the core of the conflict structure. With the three dimensional approach argued by Galtung, its inadvertently unpreventable to say that humans handle conflict resolution in a heuristic manner and not algorithmic way.

Systemic steps in conflict resolution:

- a. Understanding the problem: All conflict has origin. In the context of this paper, conflict originates from the way a speech is produced. The way the communication occurred can be non-verbal i.e. (facial expression) or verbal expression.
- b. Proffering diverse solutions: Based on Galtungs model of using a triangle, conflict resolution doesn't only involve head knowledge or brain activities. There are some inert

attributes in all humans that houses the contradictions, Attitudes and behaviours identified by Galtungs model. This inert ability is transient and cannot easily be seeing or measured. But the faculties that controls these inherent features are the mind, heart and the soul. With this, there is clarity to which faculties helps us the most in fashioning conflict resolution processes.

- c. Providing a reverse language to counter the conflict: Conflict exists out of a spoken word and to solve the conflict, another word has to be spoken to reverse/counter the conflict. Words are powerful and can either kill or make alive, just as the words fathers speaks to their children while growing up contributes to the overall making of the adulthood of the child.
- d. Testing: This involves a round table call to put into practice the perceived solution to the conflict having considered various instances and arriving at a resolute state where all parties are happy. This stage might not be automatically accepted by the other party because human behaviours can't be controlled but managed, and conflict resolution requires efficient management of non-verbal communications like facial expression, grumbling, pace at which individuals talks, pace at which they walked into the meeting, etc.
- e. Running the loop again: This involves going through the process again. This always happens when the first conflict fails. Information is gathered and a better solution is proffered as the loop runs again.

3.4. Ways Computers Handle Conflict

The conflict must have happened before and an already programmed approach for handling the conflict must also exist. In computer programming there is a particular block of code where there is possibility of having a conflict and the whole process is included in what's called a "try catch block".

The try catch block is represented in pseudocode format below:

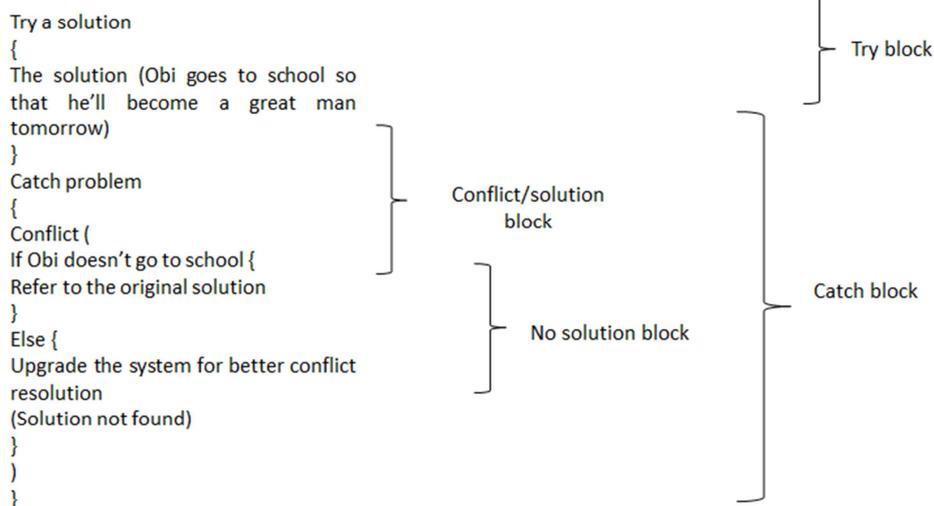


Figure 3. Try catch block representing how computers handles conflict.

The sketch above shows that the solution is inbuilt into a system, using a large repository to search through a given database for the solution of a conflict via an artificial intelligent approach. On finding no solution, the system crashes until the system is upgraded to solve the new problem.

4. Conclusion

Human speech production is still by far better than computer synthesized speech in terms of conflict resolution. The efforts by scientist to embed some human-like functions into the computer system is appreciated. They have tried in terms of static representation of speech and also to an extent, a dynamic flow through the introduction of artificial intelligence. Yet, conflict resolution is a system on its own that involves several faculties of the human being, for example the brain, heart, mind and also soul, while computer system utilizes an artificially formed brain that behaves like another without a mind of its own or soul. Furthermore, all the functions of real humans can't be in a system controlled by humans, given the result of the research, but most of the functions can be fashioned out through in-depth research and laboratory testing.

5. Recommendation

It's not entirely impossible for human-like functions to be embedded on a computer system, but as it stands now, an in-depth research should be done on the various faculties involved in conflict resolution. If those faculties can be pinned down into variables and functions, then it's possible to build artificial systems (computer system) that can be involved in real conflict resolution.

References

- [1] Beaugrande and Dressler (1992). Nigeria and the role of English language in the 21st century Retrieved from <https://eujournal.org/index.php/esj/article/download/1153/1169>.
- [2] Nhlapo, T., Arogundade, E., & Garuba, H. (2014). Things fall apart? reflections on the legacy of Chinua Achebe.
- [3] Fagyal, Z. (2001). Phonetics and speaking machines: On the mechanical simulation of human speech in the 17th century. *Historiographia linguistica*, 28(3), 289-330.
- [4] Nwakanma, I. C., Okpala, I. U., & Oluigbo I. (2014). Text – To – Speech Synthesis (TTS).
- [5] Webel, C., & Galtung, J. (Eds.). (2007). *Handbook of peace and conflict studies*. Routledge.
- [6] Ali, A. M., Asgari, S., Collier, T. C., Allen, M., Girod, L., Hudson, R. E.,... & Blumstein, D. T. (2009). An empirical study of collaborative acoustic source localization. *Journal of Signal Processing Systems*, 57(3), 415-436.
- [7] Baxter, L. A. (2006). Communication as dialogue. GJ Shepherd, J. St. John, & TG Striphos (Eds.), *Communication as—: Perspectives on theory*, 101-109.
- [8] Bergenthum, R., Desel, J., Lorenz, R., & Mauser, S. (2008). Synthesis of Petri nets from finite partial languages. *Fundamenta Informaticae*, 88(4), 437-468.
- [9] Blench, R. (2012). Research and development of Nigerian minority languages. *Advances in Minority Language Research in Nigeria*, 1, 1-15.
- [10] Cortadella, J., Kishinevsky, M., Kondratyev, A., Lavagno, L., & Yakovlev, A. (1997). Petrify: a tool for manipulating concurrent specifications and synthesis of asynchronous controllers. *IEICE Transactions on information and Systems*, 80(3), 315-325.
- [11] Cook, J. E., & Wolf, A. L. (1998). Discovering models of software processes from event-based data. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 7(3), 215-249.
- [12] Demmers, J. (2002). Diaspora and conflict: Locality, long-distance nationalism, and delocalisation of conflict dynamics. *Javnost-The Public*, 9(1), 85-96.
- [13] Dudley, H., & Tarnoczy, T. H. (1950). The speaking machine of Wolfgang von Kempelen. *The Journal of the Acoustical Society of America*, 22(2), 151-166.
- [14] Dutoit, T. (1997). *An introduction to text-to-speech synthesis (Vol. 3)*. Springer Science & Business Media.
- [15] Emejulu, O. A., Nwakanma, I. C., & Okpala, I. U. (2019). *Digital Language Mining Platform for Nigerian languages (DLMP)*.
- [16] Fagyal, Z., Kibbee, D., & Jenkins, F. (2006). *French: A linguistic introduction*. Cambridge University Press.
- [17] Henze, R., Katz, A., & Norte, E. (2000). Rethinking the concept of racial or ethnic conflict in schools: A leadership perspective. *Race Ethnicity and Education*, 3(2), 195-206.
- [18] Jabri, V. (1996). *Discourses on violence: Conflict analysis reconsidered*. Manchester University Press.
- [19] Meyer, P., Rühl, H. W., Krüger, R., Kugler, M., Vogten, L. L. M., Dirksen, A., & Belhoula, K. (1993). PHRITTS-A Text-To-Speech Synthesizer for the German Language. In *Third European Conference on Speech Communication and Technology*.
- [20] Monperrus, M., de Montauzan, M. G., Cornu, B., Marvie, R., & Rouvoy, R. (2013). Challenging Analytical Knowledge On Exception-Handling: An Empirical Study of 32 Java Software Packages (Doctoral dissertation, Laboratoire d'Informatique Fondamentale de Lille).
- [21] Ramsbotham, O., Miall, H., & Woodhouse, T. (2011). *Contemporary conflict resolution*. Polity.
- [22] Schrijver, A. (1998). *Theory of linear and integer programming*. John Wiley & Sons.
- [23] Zhao, Z. A., & Liu, H. (2011). Spectral feature selection for data mining. Chapman and Hall/CRC.
- [24] Schröder, M. (2004, June). Dimensional emotion representation as a basis for speech synthesis with non-extreme emotions. In *Tutorial and research workshop on affective dialogue systems (pp. 209-220)*. Springer, Berlin, Heidelberg.