Paper ID# 901395.PDF
"A" (Approved for Public Release, Distribution Unlimited)

AUTOMATING BATTLEFIELD EVENT REPORTING USING CONCEPTUAL SPACES AND FUZZY LOGIC FOR PASSIVE SPEECH INTERPRETATION

Katie T. McConky
Research Scientist
CUBRC
Buffalo, NY, U.S.A.
mcconky@cubrc.org

Pat McLaughlin
Director, Technology Innovation Enterprise (TIE)
Lockheed Martin IS&GS
Laguna Hills, CA, U.S.A
pat.mclaughlin@lmco.com

William Rose, Ph.D.
Project Manager
Lockheed Martin IS&GS
King of Prussia, PA, U.S.A.
william.j.rose@lmco.com

Moises Sudit, Ph. D.
Managing Director
Center for Multisource Information Fusion
Buffalo, NY, U.S.A.
sudit@cubrc.org

ABSTRACT

This research explores the feasibility of performing passive information capture on voice data in order to analyze and classify the contents of interpersonal communication. The general form of this problem is very difficult as fully automated speech understanding technology does not exist. This is further complicated by battlefield realities including: noise, jargon and unstructured speech. However, when specific topics are isolated for extraction, the challenge becomes manageable. Conceptual Spaces is used as a fusion framework to classify data passively captured by traditional speech recognition software coupled with fuzzy logic to provide matching of phonetics to jargon. Together these technologies prove to be a valuable fusion framework because of their ability to mitigate the high levels of errors inherent in speech recognition. An initial study focused on recognizing important topics in communications between commanders and field personnel amidst background chatter. Results indicate the Conceptual Spaces model is flexible enough to define "spaces" for military events, and the underlying optimization model used for classification was robust and fast enough to quickly and accurately classify the noisy scenario data. This technology enables a new and more general class of automation, permitting conversion of passive speech into structured data.

The authors gratefully acknowledge the support provided by the Defense Advanced Research Projects Agency (DARPA).

INTRODUCTION

Military operations can experience significant delays between when soldiers report a battlefield event verbally, and when the verified event report is displayed with an appropriate symbol on a command interface. Also, valuable staff hours are spent transcribing the important elements of the conversation. This delay in event reporting reduces situational awareness (SA) timeliness, impacts planning, and prevents automated fusion with other real-time data sources. In order to increase SA and reduce battlefield event reporting latency, this research examines the feasibility of automating battlefield event reporting by combining traditional speech recognition techniques on radio transmissions with Conceptual Spaces and fuzzy logic as a fusion mechanism to classify the speech data into structured reports suitable for automated distribution and action. There are three structural components of the envisioned system: 1) a fuzzy logic based speech augmentation system that uses grammars to extract and translate military specific terms based on contextual information, 2) a Conceptual Spaces (CS) inference system which computes the similarity of normalized conversations to predefined concepts, in this case, critical event reports, and 3) an automated event reporter which generates a structured report based on the CS model for consumption and automated action by electronic processing systems such as command post of the future (CPOF). CPOF uses electronic event reports to automate notification of key resources, support data mining research, and support planning such as in a medical evacuation scenario. The effectiveness of CPOF is reduced when the events must undergo the long delays often associated with manual transcription during critical events.

The Passive Information Capture and Notation (PICaN) approach uses automatic speech recognition to generate a

list of candidate utterances, fuzzy logic to integrate background information and improve speech recognition, and conceptual spaces theory to produce a fuzzy classification of the reported event and an appropriate associated icon. The PICaN approach aims to reduce reporting latency by publishing event reports in near real time and will improve event labeling accuracy through its use of conceptual spaces for event classification. This approach also benefits from requiring no additional equipment or training for soldiers in the field.

The remainder of this paper outlines the use of speech recognition to passively capture dialogue, fuzzy logic to improve the speech recognition in a military context, the use of conceptual spaces to classify speech data into events, the design and results of an experiment to test the system, and conclusions and suggestions for a full PICaN system architecture.

SPEECH RECOGNITION IN CONVERSATION

Speech recognition software falls into two general categories: speaker-dependent and -independent. Speaker dependent speech recognizers are designed for general speech but specific speakers. This software can be trained with high accuracy to correctly interpret an individual speaker's voice and works especially well for conventional English. Unfortunately, speaker dependent speech recognizers require significant amounts of time training the recognition software and work poorly in noisy environments. Speaker independent software was designed for applications that are quite common today in automated telephone interaction. These work by substantially constraining the possible word choices, often in a hierarchical structure. So in one case the vocal range is constrained, and in the other the phonetic set, but both recognition paradigms present significant constraints. In the military application envisioned, training the software to recognize each soldier in the field would not be practical nor would imposing word choice constraints. In contrast, the TOC operators are well established so training would be possible. This research concluded that the simplest approach was to have speaker independent software doing only keyword spotting for soldiers and more robust speaker dependent software at the TOC trained to the user. This works extremely well in a military environment because operators are trained to repeat all critical information, so even if the field user end is noisy, the TOC operator is quite clear and the software can be well trained. Even so, complex grammars and fuzzy contextual reasoning are required to make this structure effective.

Military dialogue contains much more varied speech than most automated systems are typically designed to handle. The highly uncontrolled military speech environment requires a large vocabulary of speech recognition terms. The large grammars required for the speaker independent speech recognition posed significant challenges and will be discussed in the following section.

In addition to challenges embedded in speech recognition, conversation has unique attributes. First and most complex is context. All conversations include hidden context which is used by the speakers to interpret information. Thus, if in a military situation someone reports their unit is under fire; the operator understands their location, the size of the unit and that fire means weapons and not wild fires because of context. Other forms of context are present in conventional conversation such as "How's the weather there?" Both parties know where "there" is. The second conversation attribute is structure. Many understanding systems take advantage of grammar formalisms to aid in understanding. Sentence structure is not a useful constraint on conversations. "Enemy fire three o'clock!" would be not recognizable by such systems but is quite clear in speech. A third conversation attribute is jargon, or micro-linguistics. These are speech constructs that are well understood by the participants, but are not standard English. In texting, this includes such "words" as u, r, lol, emoticons, etc. In the military, these are BREV-codes, such as SITREP, or pre-defined designators such as checkpoint Charlie (particular latitude and longitude).

GRAMMARS AND PERFORMANCE

Developing of custom grammars is a particularly important aspect of speech recognition for this and related environments because a significant number of important words are not English. In our battlefield example, we recognized 6 categories of necessary extensions:

1) Standard abbreviations: these are predefined jargon/terms associated with military operations such as SITREP.
2) Locally assigned names: these are terms specific to the local mission and often are otherwise nonsensical word combinations. For example, Red Charlie may refer to a specific operation phase, or Thrifty Green may be a unit designator.
3) Stop Words: these are terms with specific military contextual meaning such as "roger", "break," or "over." These terms are useful in segmentation.
4) Index identifiers: these are terms that are actually implied references, for example "SITREP" followed later by "alpha" refers to line A of a SITREP report
5) Slang and acronyms: these are abbreviations that may be pronounced as letters, such as "I E D" or letters replaced by words, such as "mm" (millimeter) maybe pronounced "mike mike." In this same category are

word substitutions for letters such as “license plate Tango, Mary, …”

6) Letter-number mixtures: these are quite common forms including AK47 and 25mm round.

Although this broad spectrum of terminology presents a challenge for grammars, in the military, these terms exist in a variety of electronic forms suitable for grammar translation.

The PICAN program used the Unisys Natural Language Speech Assistant to rapidly construct the speech recognizer grammars, which were deployed on the SRI DynaSpeak speaker-independent speech recognizer. For speaker-dependent recognition, we used the Microsoft SAPI 5.1 engine. Two difficulties were encountered in the initial speech recognition process attempts with speaker independent technology related to monolithic grammar files and free form remarks.

A large monolithic grammar file, which contained all possible combination of expected utterances, produced a high rate of substitution, deletion, and insertion errors due to self-imposed limited recognition time and high level of required backtracking. These errors were reduced by creating separate grammars for each possible battlefield event reported; essentially creating a parallel processing environment.

Initially, results were poor when attempting to recognize the longer ‘remarks’ sections of event reports in speech, which have the possibility of containing elements which fall outside the grammar, producing a high rate of false positives. This was mitigated by using a reject-word strategy and returning only recognized key-words.

Since a single grammar was created for each event, the raw speech data was simultaneously fed into parallel speech recognition software instances, with one instance for each event grammar. Running several instances of speech recognition software in parallel allowed the individual event grammars to be fully processed without timing out, and allowed the system to still respond in a near real time manner.

Even with adaptations to the grammar files, speaker-independent speech recognition data was characterized by high error rates, with both significant numbers of false positives and false negatives. False positives were words that appeared in the results that did not actually appear in the spoken scenario. Of the total speaker independent speech recognition data, approximately 65% of the data were false positives. Of all the false positives observed, approximately 10 percent were false positives with extremely high certainty. False negatives consisted of identified key words in the scenario that did not appear in the final output. Rates of false negatives were high. Up to 70 percent of keywords did not appear in the speech recognition output. The performance of the speaker independent speech recognition software is summarized in Table 1.

Table 1: Speaker Independent Speech Recognition Performance Summary

Performance	Measure
70%	**False Negatives** – percentage of keywords missed
65%	**False Positives**- percentage of keywords said to be present but are not in source
10%	**Severe False Positives**- percentage of false positives with high certainty ratings.
80%	**Probability of Detection** – probability that the correct keywords were identified

Speech recognition results contained a list of confidence-indexed key-word phrases used as input to the conceptual spaces algorithms.

CONCEPTUAL SPACES (CS)

Conceptual Spaces(CS) is a logical and mathematical construct allowing for the integration of diverse information components originally developed by Gardenfors[2]. An illustration of a CS model is the domains of weapon and number of rounds fired used to describe the concept of a small arms fire event. A small arms fire event can be carried out by handguns, semi-automatic pistols, or assault rifles which represent allowable properties of the weapon domain, and may have a number of rounds fired property of less than 600 rounds or greater than 600 rounds. For classification purposes, we have extended CS structure to include mathematical constructs for fuzzy concept similarity computations and cross domain constraints formatted into forbidden pairs[1]. The concept of a small arms fire event may be constructed to include the forbidden pairs: (handgun, >600 rounds) and (semi-automatic pistol, >600 rounds) to indicate that a valid small arms fire event concept only allows the weapon to be a handgun or semi-automatic pistol if less than 600 rounds are fired. These forbidden pairs take into account external knowledge of the cartridge capacity of each weapon type, the firing speed capabilities, and the typical length of small arms fire events. This basic framework was employed here to determine the extent to which a processed conversation string is similar to pre-defined concepts of military events.

Conceptual spaces were used to represent the CPOF events as complex multidimensional geometries. These geometries are inherently convex polytopes, so they lend themselves nicely to optimized classification algorithms

via integer mathematical models. A conceptual space consists of quality dimensions contributing to several domains that are segmented into distinct properties. An individual concept within a conceptual space consists of a set of allowed properties for each domain applicable to a concept, as well as a set of forbidden pairs that represent cross domain constraints.

Several mathematical integer programs have been suggested to optimally classify an observation using conceptual spaces [1]. The existing approaches all look to classify an object by maximizing the amount of supporting information for each possible concept. This research looked at not only maximizing the amount of supporting information, but also maximizing the amount of contradictory information as a basis for observation classification. An object classification was chosen based on the ratio of supporting to contradictory information. The concept with the highest support/contradiction ratio was chosen as the optimal concept classification.

CONCEPTUAL SPACES MATHEMATICAL MODEL

An extension of the single observation integer program proposed in [1] was chosen as the basis for the Conceptual Spaces algorithm because a certainty score and contradiction score were needed for each possible concept. An observation consists of a set of property-certainty pairs. In this experiment, property-certainty pairs were obtained from the speech recognition output. Two integer programming problems were formulated for each concept, a certainty model and a contradiction model, and together they were used to optimally decide what concept an observation belongs to. The certainty model follows:

D = Set of Domains Considered

P_k = Set of Allowed Properties of domain k

I = Set of mutually exclusive (i, j) and (i', j')
where $i \in D$ and $i' \in D$, $i \neq i'$ and $j \in P_i$, $j' \in P_{i'}$

p^j = importance of domain j to the concept

l_{ij} = salience of property i to domain j

s_{ij} = Similarity of observation to property i from domain j

$$x_{ij} = \begin{cases} 1 \text{ if property i of Domain j is considered} \\ 0 \text{ otherwise} \end{cases}$$

$$Max \sum_i \sum_j s_{ij} l_{ij} p^j x_{ij} \quad (1)$$

$$\text{st.:} \sum_{i \in P_j} x_{ij} = 1 \qquad \forall j \in D \quad (2)$$

$$x_{ij} + x_{i'j'} \leq 1 \quad \forall \{(i, j), (i', j')\} \in I \quad (3)$$

$$x_{ij} = 0 \text{ or } 1 \quad \forall i, j \quad (4)$$

This integer program was run once for each concept in order to obtain a concept certainty score, or support value. The objective function, Equation 1, maximizes the similarity over the set of all property certainty pairs to the particular concept in question. For each model the set of domains (D) and allowed properties (P_k) are changed to correspond to the appropriate concept. Equation 2 says at least one property must be present from each domain. Equation 3 handles the cross-domain property constraints. The normalized results of this model provide a certainty score for each concept. This certainty score represents the amount of supporting evidence the observation provides for each concept. A second integer program was run to maximize the contradictory information provided by the observation for each concept. The model to maximize contradictory information for each concept follows:

D = Set of Domains Considered

P'_k = Set of Dis - Allowed Properties of domain k

s_{ij} = Similarity of observation to property i from domain j

$$x_{ij} = \begin{cases} 1 \text{ if property i of Domain j is considered} \\ 0 \text{ otherwise} \end{cases}$$

$$Max \sum_i \sum_j s_{ij} x_{ij} \quad (5)$$

$$\text{st.:} \sum_{i \in P'_j} x_{ij} = 1 \qquad \forall j \in D \quad (6)$$

$$x_{ij} = 0 \text{ or } 1 \quad \forall i, j \quad (7)$$

The objective function, Equation 5, maximizes the similarity over the set of all property certainty pairs to the particular concept in question. For each model the set of domains (D) correspond to the original concept domains, and the dis-allowed properties (P'_k) are those properties for each domain that were not allowed in the original concept. Equation 6 says at least one dis-allowed property must be present from each domain. The normalized results of this model provide a contradiction score for each concept. The contradiction score represents the amount of contradictory evidence the observation provides for each concept.

The final classification is made by choosing the concept with the highest certainty/contradiction ratio. In the event of a tie, the certainty score is used as a tie-breaker.

CONCEPTUAL SPACES FOR MILITARY EVENTS

The set of domains and properties used in this experiment were developed in alignment with the speech recognizer grammars. The domains were created around commonly discussed critical event components and included: Weapons, Events, Indicators & Equipment, Chemical Indicators, Location Descriptors, Sounds, Type of IED, Wounds, Caliber, Damage, and Device Initiation. Common words and phrases associated with these

domains were elicited from subject matter experts (SME) to form the properties for each domain. The mapping of the domains to applicable concepts can be seen in Table 2.

Table 2: Domain to Concept Mapping

Domain	Grenade	IED	Ins. Vehicle	Mine	Mortar	PBIED	RPG	SAF	Sniper	VBIED
Caliber		L			L			M	L	
Chem Indicators		L								L
Damage	L	L		M	L		L			L
Device Initiation		L		L		L				L
Equipment		L	H	L		H			L	
Events	H	H		M	H		H	H	H	M
Indicators		L	H	M	L	M	M			L
Local Descr.			M		L	L			M	L
Sounds		M		M		L		L	L	M
Type of IED		L								L
Weapon	H	H	M	H	H	H	H	H	H	H
Wounds	L	L				L	L	L	L	

A value of H means the domain has a high importance to the concept, and M or L indicates medium or low importance respectively. These fuzzy values correspond to the relative importance of the domains to a specific concept, represented by p^j values in the mathematical model. As the table indicates, concepts were partially defined by the domains they contained. An additional concept component was the allowed properties for each of their applicable domains. An example concept definition is provided for the IED concept in Table 3; a similar definition table was developed for each event.

EXPERIMENT

Using military subject matter experts, a training and test set of ten possible attack scenarios were created along with routine background "chatter". The following event types were chosen: grenade, improvised explosive device (IED), person-borne IED, vehicle-borne IED, insurgent vehicle, mine, mortar, small-arms fire, sniper, and rocket-propelled grenade (RPG).The training set consisted of one scenario for each of the events, while the test set consisted of two additional scenarios for each event, providing a total of 30 scenarios.

In addition, two non-critical event scenarios were also created to be used as negative outlying test variables. All scenarios varied in their sequence of reported event characteristics, some followed prescribed reporting structures as expected by operations, but others allowed for completely unstructured responses as sometimes happens during crisis or when events are incompletely formed. For simplicity in this initial experiment, the scenarios were constrained to include only one incident.

Table 3: Domains and Allowed Properties for IED Concept

Domain	Salience	Allowed Properties
Device Initiation	.02	Cell phone, LRCP, Motorola, nokia
Chem Indicators	.01	Blistering, chlorine, irritation, phosphorus, white phosphorus
Weapon	.4	IED, road side IED, unexploded IED, UXO
Sounds	.1	Exploded, explosion
Events	.4	IED exploded, IED went off, went off
Type of IED	.02	Artillery shell, EFP, MGM, propane tank, UBE
Equipment	.01	Shovel, shovels
Indicators	.01	Digging, disturbed earth, exposed wires, trash
Damage	.01	(Bradley|vehicle| equipment) & (damaged| destroyed|hit)
Caliber	.01	155, 60, 80, 300, 120
Wounds	.01	Shrapnel

Training and test scenarios included conversations between the unit commander and a Tactical Operations Center analyst. Labeled scripts were created to represent the conversations and were recorded into pulse-code modulation (PCM) based audio files sampled at 8 Khz, 16-bit mono. Utterance length varied from single word, i.e. "roger", to multiple word remarks containing close to 40 words.

The training set of scenarios was used to tune the speech recognizer grammars and parameters. The tuning consisted of processing the speech files through the recognition software, passing the results to the conceptual spaces algorithm, and adjusting both the speech recognition grammars and the conceptual spaces event concept definitions in an attempt to improve performance. After several iterations of tuning, the test scenarios were passed through both systems with no further performance tuning. The two non-event scenarios were also processed after initial training was completed.

RESULTS

The CS models performed very well, even when the speech recognition software had high error rates. Speaker independent systems generated high error rates, while speaker dependent systems had lower error rates. The former case was used to illustrate the systems

effectiveness even with very poor speech recognition. The latter case is more realistic since TOC operators can easily afford the application training time.

The results of the speech recognition output included files for 20 event scenarios and two non-event scenarios, with a confidence score for each recognized utterance. Each of the 22 test scenarios was run through the conceptual spaces algorithm. The output of the algorithm was an ordered list of event-certainty pairs, with the most likely event listed first. These ordered lists were analyzed to determine the overall performance of the conceptual spaces algorithm for both event classification and event/non-event recognition.

The research team investigated the performance of speaker dependent speech recognition applied only to the TOC analyst. The portions of the scenario files corresponding to TOC speech were analyzed with speaker dependent recognition software trained on the TOC speaker. The speaker dependent data produced near 100% recognition of key words, corresponding to roughly 100% probability of detection and 0% probability of false negatives. These superior speech recognition results were run through the conceptual spaces algorithm leading to 90% correct classification of events. The results of the processing of speaker dependent speech are summarized in Table 4.

Table 4: Conceptual Spaces Performance Summary For Speaker Dependent Recognition

Performance	Measure
90%	Correct Classification
10%	Incorrect Classification
0%	False Negatives

Results for the speaker independent speech recognition data were also respectable despite the inherent high error rates of the recognition system. Of the 20 event scenarios 70% of the events were classified correctly. A correct classification was determined to be either the correct event with the highest conceptual spaces certainty or a similar event containing the highest certainty with the correct event still in the top three. For example, a sniper event scenario classified as a small arms fire scenario would be considered a correct classification if sniper was also in the top three.

Of the 20 event scenarios 30% of the events were miss-classified, but the certainty level of the correct event label was still greater than the non-event threshold. There was a zero percent false negative rate, meaning no events were miss-classified as non-events.

Of the two non-event scenarios, only one was correctly classified as a non-event, leading to a 50% false alarm rate. The conceptual spaces classification results for the speaker-independent speech recognition are summarized in Table 5.

Table 5: Conceptual Spaces Performance Summary For Speaker Independent Recognition

Performance	Measure
70%	Correct Classification
30%	Incorrect Classification
50%	False Alarm
0%	False Negatives

DISCUSSION OF RESULTS

The results indicate that the conceptual spaces algorithm has a desired bias towards preventing false negatives. None of the 20 real event scenarios were miss-classified as non-events, and the speaker dependent speech recognition results allowed for 90% correct event classification.

The results of the speaker independent speech recognition, despite significant efforts to increase performance by adjusting grammar files and software parameters, were poor. Speech recognition results were characterized by high rates of false positives and false negatives, leading to important words dropped from the output and a large set of unexpected words included in the output. Poor speech recognition may be due to a number of variables, including less-than mature speech recognizer models, sub-optimized recognizer parameters, and/or information overload, i.e. the 2000 millisecond recognition window was too small to capture the significant key word(s) in longer utterances.

The poor speaker independent speech recognition data challenged the conceptual spaces algorithm: The high false negative rate of keywords left conceptual spaces with few event identifying words to process. Additionally, the false positives with high certainty provided a significant amount of misleading evidence. Despite these data shortfalls conceptual spaces was still able to correctly classify 70% of the event scenarios. This suggests two things: (1) the conceptual spaces framework can be successfully used to describe military events, and (2) the conceptual spaces classification algorithm performs well even with data with high uncertainty and poor quality. Ideally, once the event is correctly identified, the system would match it with an accepted military icon/symbol and transmit a structured report, along with an uncertainty factor, to a receiving system such as CPOF.

POSSIBILITIES FOR FUTURE RESEARCH

This initial research study has provided valuable insight into future research applications. Some future extensions to this research include:

1) Non-Real-Time Systems. All of the systems we experimented with were designed for real-time

responses. This constrained their reasoning process. When the extended grammars were included, the systems did poorly since results had time constraints. In our envisioned application of a passive background system, accuracy is more important than time. A few extra cycles improving accuracy are well worth modest delays.

2) Automated Grammar Ingest: most of our grammar structure was developed manually. However, much of the proper military report data exists to allow automated ingest.

3) Handling of Multiple Event Scenarios: communications where multiple events are reported by a single source need to be handled. For example, an IED detonation may be followed by small arms fire. The classification algorithm needs to be augmented in order to classify the simultaneous events properly.

CONCLUSIONS

An experiment was completed to examine the feasibility of automating battlefield event reporting and the conversion of attack events into a standard reporting format and military symbol. By combining speech recognition, fuzzy logic and conceptual spaces algorithms, initial results are excellent when we consider the operator with speaker dependent processing, and are promising in the much more hostile environment associated with noise and speaker independent processing. In order to examine possible methods of increasing the accuracy of event detection and mapping, the following algorithms have been briefly tested as a forward look to an eventual system:

The use of speaker-dependent speech recognition to enhance event identification accuracy through the analysis of clarification speech offered by the TOC analyst to the field commander, since almost each phrase communicated by the field commander to the TOC analyst is confirmed in full utterances by the analyst in a controlled environment;

- Normalization, through tested and well known phonetic processing technologies, of standard military acronyms, grid locations, call signs and other commonly used abbreviated phrases, into fully formed phrases;
- Automatic identification of changes in tactics related to events and sub-events by the extraction of word associations;
- Fusion of conversationally derived event data with other sources including sensors and citizenry reporting.

A proposed architecture taking into account these suggestions and other system enhancements is seen in Figure 1. This proposed system architecture provides a system to convert ad hoc contacts/reports into validated actionable information streams. Extensions include fusing citizenry, legacy military patrol, and sensor data-reports, machine learning algorithms for novel event characteristics, and entity extraction of speaker dependent and independent speech recognition to enhance event classification and reporting.

REFERENCES

[1] Holender, M., Sudit, M., Nagi, R., Rickard, T., *Information Fusion Using Conceptual Spaces: Mathematical Programming Models and Methods*, 10th International Conference on Information Fusion, Quebec City, Canada July 2007, pp. 1-8.

[2] Gardenfors, P., *Conceptual Spaces: The Geometry of Thought*. 2000, Cambridge, MA: MIT Press.

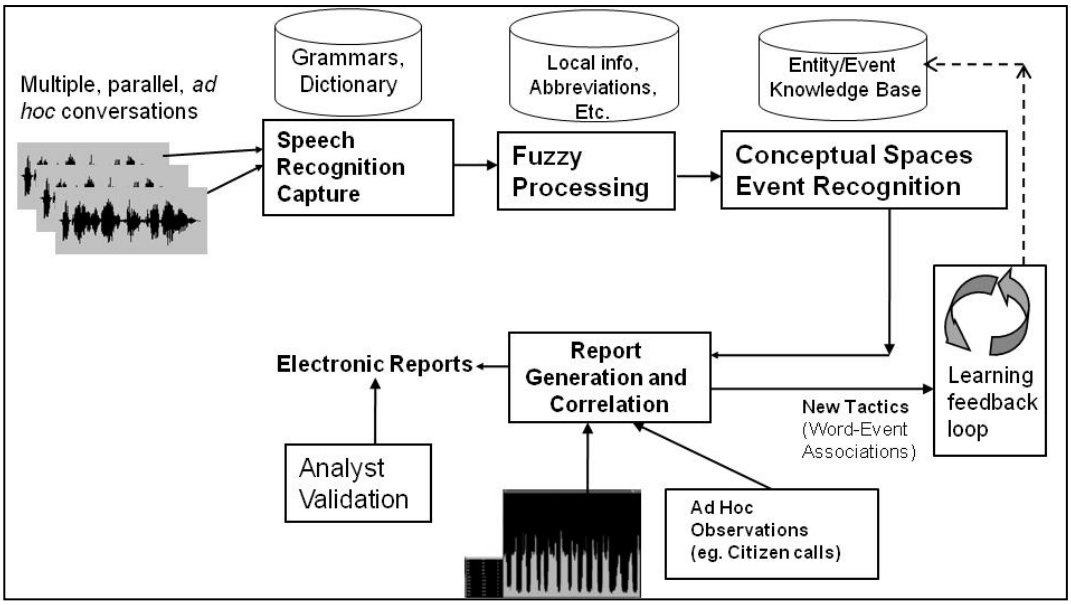

Figure 1: Proposed PICaN System Architecture