nemesysco
voice analysis technologies

Layered Voice Analysis (LVA)

Technology White Paper

Introduction

Mankind, throughout history, has sought ways and means of separating truth from deception. From ancient techniques such as hot irons, the wheel and burning at the stake, to modern 20th century polygraph, truth serum and voice analysis.

Layered Voice Analysis Technology (LVA) was invented in Israel in 1997, originally for security purposes. Today, Nemesysco technologies provide voice analysis solutions for both defense and civilian markets and based on emotion detection through voice analysis, optimize the processes of truth verification, expose malicious intentions, provide risk assessments and reveal emotions.

LVA is comprised of a set of unique signal processing algorithms that identify different types of stress, cognitive processes and emotional reactions. 129 vocal parameters are utilized in order to detect and measure minute, involuntary changes in the speech waveform and create the foundation for identifying the speaker's emotional profile.

Enabling analysis of conversations in real-time or offline (from recorded files) and providing quick truth verification, LVA makes investigators' work more efficient and improves decision making processes.

LVA Technology

LVA technology identifies a subject's state-of-mind by analyzing key vocal properties in his/her speech

- Identifies a subject's state-of-mind by analyzing key vocal properties in his/her speech
- Identifies various types of stress, cognitive processes and emotional reactions
- Creates an "emotional signature" of an individual's speech at a given moment
- Detects deceptive motivation, criminal intention, and general credibility by identifying key emotional signatures reflected in the voice

LVA and Lie Detection

Since lying is not associated with a single unified set of feelings that can be measured, there can not be such a thing as a "real" lie detector. Lying is a result of a deep logical process that has a particular intention behind it.

Some people will lie to protect themselves from harm, while others will lie to gain profit, or even just to make a joke. Due to the various reasons for lying, it is difficult to ascertain a fixed set of characteristics (physiological or psychological) that differentiate lies from truth.

LVA is capable of detecting the *intention* behind spoken words. Detecting this intention allows to identify and reveal the lie itself. This unique functionality brings LVA as close as possible to being an emotion detector, used for lie detection purposes.

The Human Speaking Mechanism

The human speaking mechanism is one of the most complex processes the human body is capable of, due to the number of muscles and physical apparatus involved, and the ways in which they need to be synchronized in perfect timing.

Initially, the brain comprehends a given situation and the possible implications of whatever will be said. Then, when a person decides to speak, air is pushed upward from the lungs into the vocal cords. This causes the vocal cords to vibrate at a specific frequency and produce sound. The vibrated air continues to flow up toward the mouth where it is manipulated by the tongue, teeth and lips to produce sound streams which we interpret as words or phrases.

LVA technology ignores what a subject is saying and focuses only on changes in the brain activity that are reflected in the voice

The brain closely monitors all of these procedures, ensuring that the sound emitted is the one that was intended, is intelligible and is at a volume that can be heard by the intended listener. Due to this constant cerebral monitoring, every "event" that passes through the brain leaves a trace on the speech flow. LVA technology ignores what a subject is saying (i.e., the specific content) and focuses only on changes in the brain activity that are reflected in the voice.

Human Voice and Frequencies

The human voice is comprised of a combination of many frequencies, sorted in several groups known as formants. The formants are typically a duplication of one another in different amplitudes, and differ from one person to the other. The normal range of frequencies in human males is between 100Hz to 3 KHz. The female voice is typically higher and ranges between 200Hz and 6 KHz.

Technology

LVA technology is based on the mathematical modeling of voice variations and irregularities, researched and found to correlate with different types of brain activities, emotional states and physiological conditions.

LVA technology affords a better understanding of a subject's mental state and emotional makeup during the time in which he or she is speaking. The technology identifies various types of stress, cognitive processes and emotional reactions. This information provides insights into the way the person thinks; what troubles him; what excites him; what portions of his response he is uncertain about; what topics require more of his attention, and what areas appear to be sensitive issues for him. In addition, LVA allows for the exploration of several levels of conscious and unconscious thoughts and feelings. This exploration can reveal additional layers of information that would otherwise not be available.

LVA uses a patented and unique technology to detect "traces" of brain activity using the voice as a medium. The technology is based on the idea that changes in cortical perception and interpretation of events manifest themselves in the vocal waveform during speech. By utilizing a wide range spectrum analysis to detect minute, involuntary changes in the speech waveform, LVA can detect variations in brain activity and classify them in terms of stress, excitement, deception, and varying emotional states, accordingly. Stress ("fight or flight" paradigm) is only one component of the overall emotional structure that is detected.

The core of the LVA technology derives from information generated by algorithms that examine the minute changes within the Relative High Frequency Range (RHFR) and the Relative Low Frequency Range (RLFR). The multitude of permutations between these two ranges of voice is what we see, hear and analyze.
Relative High Frequency Range (RHFR): Examples of this would be seen in high states of excitement and emotional arousal.
Relative Low Frequency Range (RLFR): Examples of this would be seen in states of stress, thinking and other cognitive processes like imagery.

Features and Benefits

The Layers of Voice Analysis

The term "Layered Analysis" indicates that LVA uses several different approaches and analytical models to reach its conclusions, based on a "layer upon layer" structure. The LVA analysis process is comprised of several steps (some of which are dependent on the desired functionality of the system):

1. **The Segmentation Process Layer.** The voice is a very sensitive type of input, highly influenced by background noise. Identifying the relevant portions of the voice data stream as actually containing human voice is crucial for the correct analysis process. Adding too much noise to the analyzed segment will render meaningless results, while leaving a significant part unanalyzed may miss important indications. The collection of the "as much as possible" exact portion of the voice is not a trivial task, and LVA uses several proprietary techniques to achieve this optimal result.

2. **The Secondary Screening Process Layer.** Once a voice segment was marked for analysis, it will undergo a secondary screening process of normalization and re-selection of the most relevant portions that are suitable for analysis. Parts of the human language components cannot be analyzed and removing these unwanted voice components is necessary for improving the accuracy of the analysis.

3. **The Time Domain Analysis Layer.** During this phase, the exact "analysis worthy" portions are scanned by the technology in order to identify the primary "emotional indicators". These are tiny sequences of samples that create inaudible (and therefore, not controlled) deviations from the normal speech patterns of the speaker, in a given time. These sequences are measured and analyzed using both quantifiable measurements and a statistical set of rules, to create the second level of vocal parameters LVA uses for its analysis (e.g. the SPT, SPJ and other parameters, as listed below).

    The exact "analysis worthy" portions are scanned by the technology in order to identify the primary "emotional indicators"

4. **The Frequency Domain Analysis Layer.** At this stage, a very specific range of frequencies of the selected voice portion is further examined over time, using a unique method of FFT (Fast Fourier Transform). The frequency modulation is examined for the presence of unique components, stability of the structure, and other commercially-confidential indicators.
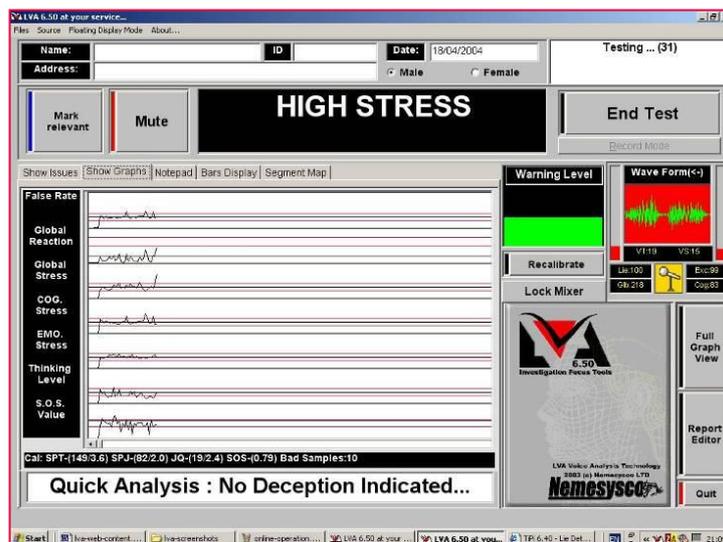
5. **The Calibration Phase Layer.** Depending on the mode of operation, LVA calculates a 3-dimensional baseline of the "Emotion Free" state (subjective normal emotional state of the tested party at the time of the test). This is performed by using either the first few voice segments analyzed, or, in case of pre-recorded material analysis, the most suitable voice segments found in the entire recording. The baseline is comprised of a type of an "average" value for most of the vocal parameters, the sensitivity of these parameters (in terms of high and low values) and their stability over time. Using this baseline and depending on the type of utilization, LVA will compute a unique "Risk Formula" which is by nature a subjective measurement.

6. **The Primary Analysis Layer**. Once the calibration baseline is calculated, every set of vocal parameters from every voice segment is compared to the baseline, taking into account the sensitivity of the specific parameter. The output of this phase reveals the level of excitement, confusion/certainty, stress (Fight or Flight), concentration, anticipation, hesitation, embarrassment and mental effort invested in the relevant statement.

7. **The "Risk" Calculation Layer.** Using the subject's unique Risk formula calculated in the calibration phase, the primary analysis indicators are used to calculate the overall deviation from the truthful baseline. The higher the deviation is, the higher the likelihood of deception or falsehood.

8. **The Lie Probability Layer.** The same set of basic and calculated emotion parameters are now measured against a statistical formula encapsulating more than 10 years of research, to reach an objective statistical measurement of the chance of falsehood.

9. **The "Deception Pattern" Detection Layer.** At this point, the whole emotional structure is put once more to the test, against pre-defined sets of emotional structures, known to be indicative (in various degrees of probabilities) of different types of deceptions (i.e. derived from different motivations).

10. **The Emotion Detection Layer.** If needed in a specific type of analysis (for example, in Nemesysco's QA5 SDK for call center vendors), the vocal parameters are used to generate the "bag of emotions" output, which includes: anger, happiness, sadness, stress, concentration, excitement, confusion, hesitation, anticipation and embarrassment, all in a normalized range of values between 0 to 30.

> *The vocal parameters are used to generate the "bag of emotions" output, which includes: Anger, Happiness, Sadness, Stress, Concentration, Excitement, Confusion, Hesitation, Anticipation and Embarrassment*

11. **The LioNet™ Technology Layer**. A trainable heuristic decision engine, designed to further improve the general analysis, as well as to learn to identify, based on experience and user feedback, new types of emotional structures, indicative of various emotional states, e.g: "ready to buy", or "cancel order", that may be suitable for call center applications.

12. **The Final Analysis Layer**. Depending on the type of utilization, LVA examines all the analysis phase results, in order to reach a final textual analysis for the analyzed segment, based on an internal hierarchy of rules. If no unique indication is found, the analysis will be "Truth". If only some excitement is detected, the system will render the "Excitement" analysis, and so on. LVA has 5 levels of falsehood analysis: "Slight Risk", "Suspected", "Inaccurate", "Probable False" and "High Risk". The determination of falsehood, if required in the desired type of LVA utilization, will only be rendered if sufficient data (both the Risk Level and Lie Probability) supports this analysis.

13. **The Overall Analysis Layer** (The "Emotional Signature" of a complete session). Once the conversation or test session is over, the technology will create a vector of numbers summarizing all the emotional activity detected across the voice segments. This vector can then be processed by the LioNet engine again, to determine a complete session classification according to the user's preferences or the specific system's needs.

The Fundamental Analysis Groups

In its fundamental layers, the technology identifies several key indications from the wave form, each indicative of a different type of brain activity or cognitive process. **"T's"** for example, are indications of a local high - frequency intervention on the wave form, typically associated with emotional reaction. **"J's"** are occurrences

> *In its fundamental layers, the technology identifies several key indications on the wave form, each indicative of a different type of brain activity or a cognitive process*

where a relatively low-frequency is superimposed on the main (fundamental) frequency (Fo) and are associated with logical processes. **"A's"**, **"X's"** and **"F's"** are more complicated sequences and their exact description is a commercially-confidential know-how of Nemesysco. These fundamental indications are then used to calculate 2 arrays of structures and 51 different computed vocal/emotional parameters, typically categorized into six groups:

1. **The Emotional Group** – used for rendering of Emotional Stress, Excitement Level, Happy and Sad Analyses, Anger (partly) and General Session Atmosphere.

2. **The Logical Group -** used for rendering of Cognitive Stress, Confusion/Certainty Levels, Mental Effort, Imagination and Hesitation Analyses (also partly used for Anger Detection).

3. **The Energy Group** – used to provide supporting information for some of the emotional analyses and stands as a valuable set of indications by itself.

4. **The Stress Related Group** – used to generate the General Stress Analysis (Fight or Flight detection), and other indications used internally to support the Lie Probability formula and the detection of Clinical Stress.

5. **The Stability Group** – used to calculate the concentration, embarrassment, SOS (say-or-stop) indications (hesitation), anticipation and arousal.

6. **Special Indicators Group** – used in different layers of the analysis process, typically supporting the detections made by one of the other groups.
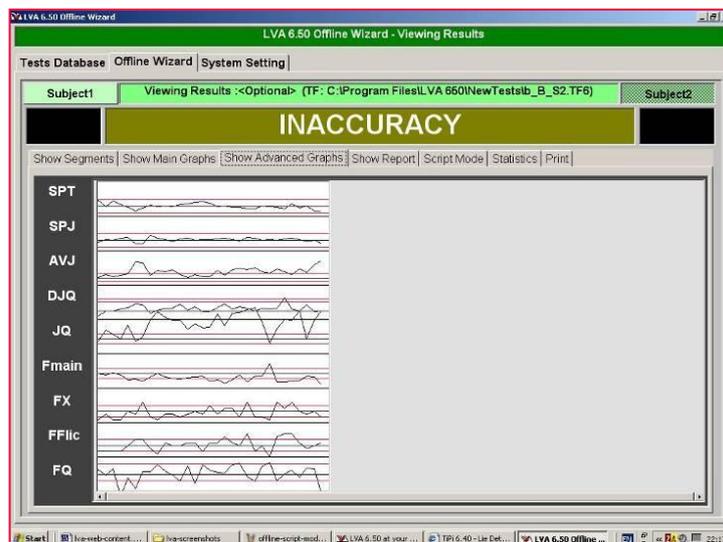
Basic Parameters and Definitions (Partial list)

SPT: This is a basic parameter, obtained from a mathematical sum based on high-frequencies superimposed on the wave, to determine the emotional level. This basic parameter and its calibration baseline are used to calculate the Emotional Stress Indications.

SPJ: This is a basic parameter, based on the low-frequencies superimposed on the wave, to determine cognitive conflicts and general cognitive activity (the whole process of knowing, thinking, judging, and learning). This basic parameter and its calibration baseline are used to calculate the Cognitive Stress and related analysis.

JQ: This parameter is associated with the global stress level (also known as the Fight or Flight syndrome). This basic parameter and its calibration baseline are used to calculate the General Stress level and related analysis.

AVJ*: Indicates the mental effort being put into what the tested party is saying. This is a measure of how intense was the cognitive effort. AVJ is used to calculate the "Thinking Level" outputs.

SOS: "Say or Stop". This parameter measures the willingness (excited to speak) or unwillingness (not excited to speak) of a person to discuss an issue and is often a signal to peek into other issues.

Fmain: This parameter represents the relative contribution to the waveform of the most significant frequency in the analyzed spectrum. This parameter focuses on concentration, tension and rejection.

FQ: This parameter is associated with negative feelings (guilt knowledge). It is a reflection of the rapidity and extremity of frequency changes in that segment of time.

SUB Cog: Subconscious cognition levels can represent the secondary cognition levels behind what is being said.
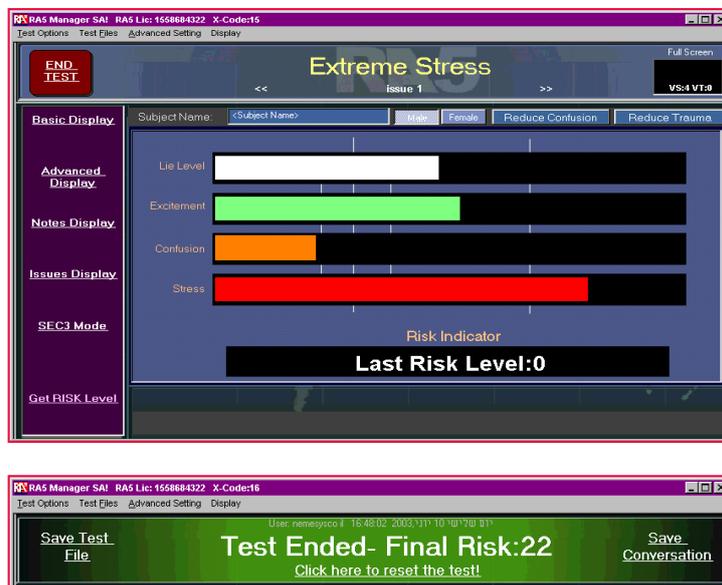
SUB Emo: The subconscious and/or profound emotional parameter is associated with the emotional part of the subconscious that clues to what emotional issues are not reflected in the conversation or are motivating the overall responses to a question.

"Risk Level": Known as subjective lie stress. It is the result of the individually unique Risk formula created using the calibration baseline. This reading measures the individual and compares these measures against his own previous reactions, to take into the account his overall state of mind and emotional set.

False Probability: Known as the objective Lie Detection formula, this statistical formula uses both the basic parameters and the calculated values, to generate a statistical lie probability value. It measures the subject against the normative value (pre-programmed parameters) of all people at all times.

> Note: Risk Level and False Probability work in conjunction with one another. The greater the propensity of the segment in being false is when both are high. If both are low then the greater the chance is for truth.

The statistical formula using both the basic parameters and the calculated values, to generate a statistical lie probability value

Cognitive Stress: Defined as two or more NON-complimenting cognitive (thinking) processes that are present in the brain at the same time.

Emotional Stress: This reads the overall emotional activity involved in what an individual is saying, typically indicative of a positive reaction.

Global Stress: Based on the normalized result of JQ, it takes into account the parameters associated with negative physical arousal, fear and alertness.

Anticipation: Anxiety that is felt in anticipation to get to a certain question or assertion of a reception of the provided answer by the examiner (convincing the examiner).

Global Reaction: Calculates the total sum of the different stress indicators in absolute values.

Security Use vs. Personal Use (the conversation atmosphere)

Security users require a tool that will provide them with just the right amount of information necessary for the case in question. Any non-relevant data can obscure what is essential. LVA technology enables the identification of relevant pieces of information quickly, and provides critical data regarding risk of deception. This allows the user to focus investigation time on potential leads and reduce the time spent on non-essential or irrelevant data.

While Nemesysco's personal-use products provide more generalized data, LVA has a dynamic range of sensitivities that enables users to zero-in on those emotional indicators that are absolutely relevant. In addition, LVA is equipped with tools that allow exploring a flagged piece of information more thoroughly. LVA technology can be utilized both in real-time (for a general overview of any subject/case/suspect/witness), and in offline mode (for a more in-depth exploration), using recorded data from almost any source.

Summary

LVA technology enables quick and effective decision making processes, based on any available audio data. LVA has the ability to identify various types of stress, cognitive processes and emotional reactions, which combine to build the complete "emotional structure" of an individual at the time of voice capture. The technology uses a series of complex signal processing algorithms to identify different emotional levels and make determinations regarding an individual's veracity, criminal intention and general credibility.