



(S) Technology to Help Transcribers

FROM: SIGINT Communications
Unknown
Run Date: 10/14/2003

(S//SI) As we all know, the volume of communications around the world has been increasing exponentially in recent years. Many SID offices have searched for new ways - using the latest technology - to find a means for keeping pace with this explosion. There are many tools being developed to help us better separate the wheat from the chaff, allowing our limited pool of analysts to concentrate on what is truly important. One such effort is FORESTPROWL, an R&D effort using Phonetic Indexing technology from Nexidia (formerly Fast-Talk).

(S//SI) With their corps of highly-qualified Spanish transcribers dwindling steadily, more traffic than they could process coming in, and a target that didn't lend itself to processing by other tools, the Office of International Crime and Narcotics (ICN) went looking for help. ICN eventually undertook a proof-of-concept project using a commercial software product developed by a research consortium to identify high-potential collected voice files. ICN contracted with ACS Defense to test the software under SIGINT conditions and to prove that it could identify specified words in a body of voice files.

(S//SI) The ultimate goal is to build a capability to help prioritize voice files for transcription based on analyst-built criteria. The FORESTPROWL software will provide the phoneme (unit of speech) "diary" of the content, and a separate engine will search and score the files for prioritization based on analytic input. Not only will it point to the highest-potential files, but it will also be a way of capturing senior analytic knowledge, since expert analysts know best what nuggets to search for and will use that knowledge to drive the selection/scoring engines. Another benefit of this effort is that junior language analysts will be more productive while in a learning mode: they will be processing the highest potential traffic rather than wading through myriad files to find the best information.

(S//SI) Encouraged by the results of the proof-of-concept effort, ICN is moving forward, using DCIP (The Defense Counterdrug Intelligence Program) money to build a complete tool, with the Nexidia engine as the centerpiece. A tweaking of the acoustical model is already underway. ICN is working with R6 to build a scoring model, and with NUCLEON (multimedia database) developers to integrate FORESTPROWL into the voice processing architecture. At the same time, the Nexidia is working on developing phoneme patterns for other languages. They have recently completed an Arabic model and are working on other Asian languages at this time. Although developed in ICN, it is hoped that the FORESTPROWL effort will prove useful to language analysts throughout the SIGINT enterprise.

(U//FOUO) POCS:

- [REDACTED] Technical Director, Office of International Crime and Narcotics, [REDACTED]
- [REDACTED] ACS Project Lead, [REDACTED]
- [REDACTED] SAIC Technical Lead, [REDACTED]
- [REDACTED] SAIC Project Lead, [REDACTED]
- [REDACTED] CACI Project Lead, [REDACTED]

"(U//FOUO) SIDtoday articles may not be republished or reposted outside NSANet without the consent of S0121 ([DL sid comms](#))."

