SAND2017-9482C

Process Map Development for Speaker Recognition

David Marks

Sandia National Laboratories

OSAC – Speaker Recognition Subcommittee

Introduction

Scientific Working Group for Speaker Recognitions (SWG-Speaker)

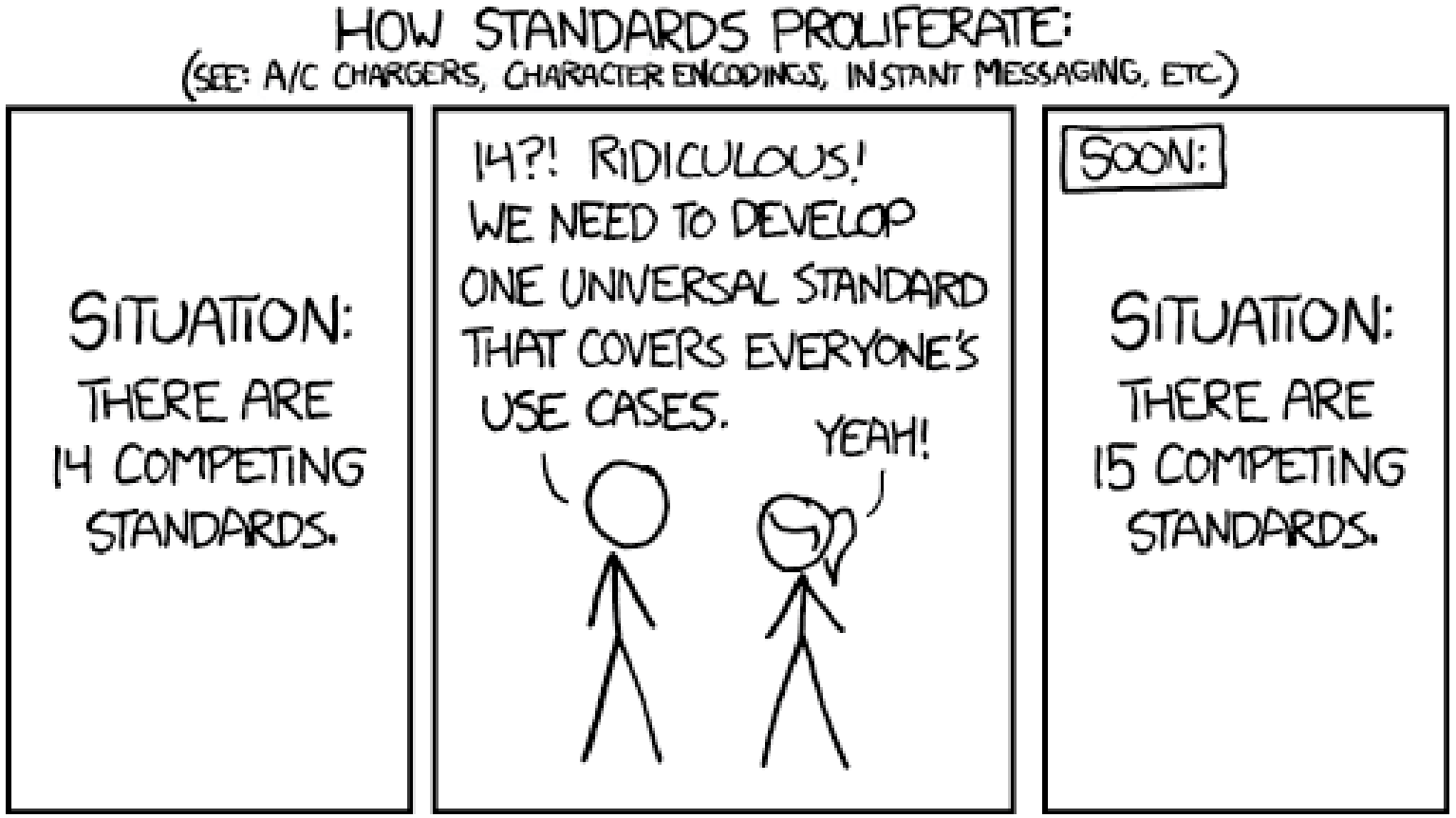
Organization of Scientific Area Committees (OSAC)

- Speaker Recognition Subcommittee (OSAC-SR)
- https://www.nist.gov/topics/forensic-science/organization-scientific-area-committees-osac
- https://www.nist.gov/topics/forensic-science/speaker-recognition-subcommittee

What is a Process Map?

Lead to ~~Standards~~, ~~Best Practices~~, ~~Guidelines~~, Common Methods

Standards



Source: https://xkcd.com/927/

History: SWG-Speaker

SWG-Speaker / Best Practices task group

Initial Efforts to develop best practices

- Terminology
- Process differences
- Agency-unique steps
- Trying to do “should-be”

Submission of straw document

- Split into 4 parts
- Split again into 2 more parts

Struggled to gain traction

History: OSAC-SR

OSAC “diversion” getting reestablished

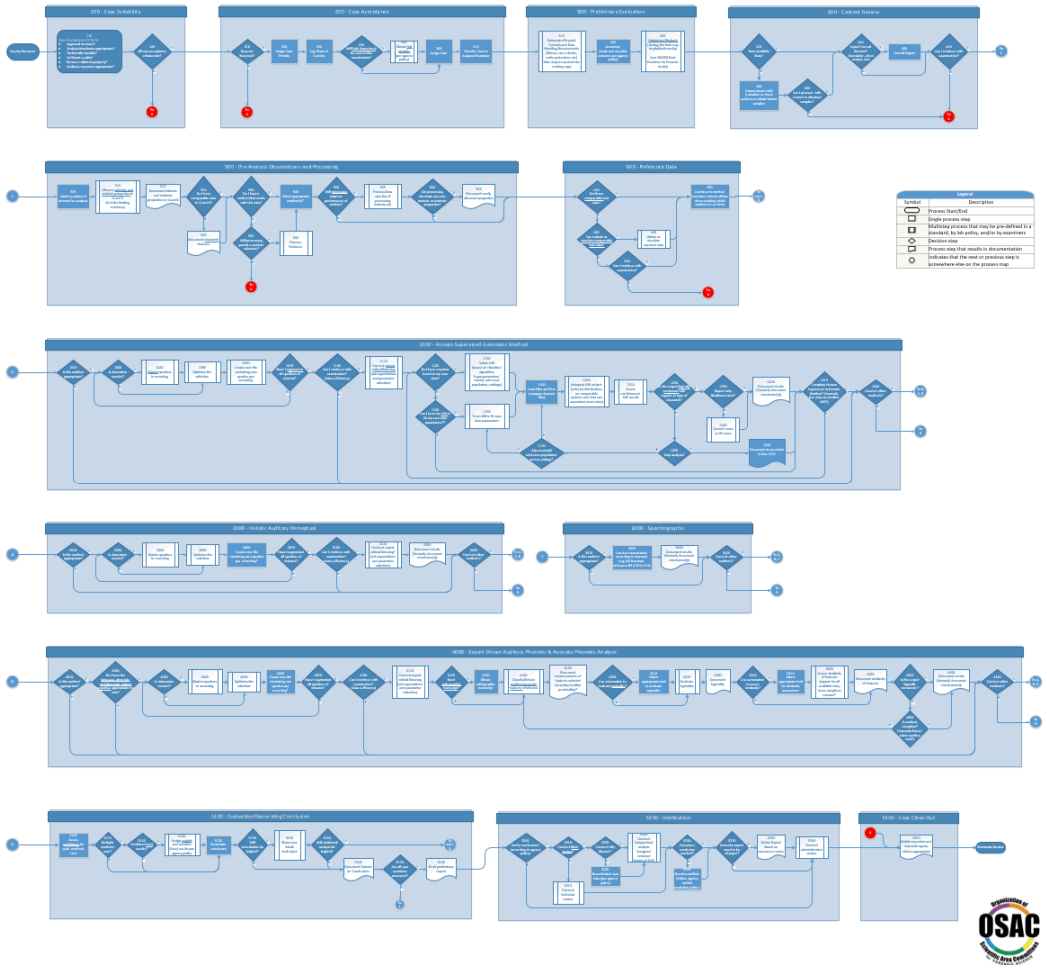
Best Practice work still floundered

Process Map exercise

- Participants
  - Practitioners, developers, researchers
  - Multiple agencies
  - Multiple countries
- Captured “as-is” processes over 3 full days (80%)
- Brainstorming – no judgement, captured multiple methods
- Huge unwieldy diagram

Issues became clearer

Initial Process Map



Process Map Refinement

Iterations/editing

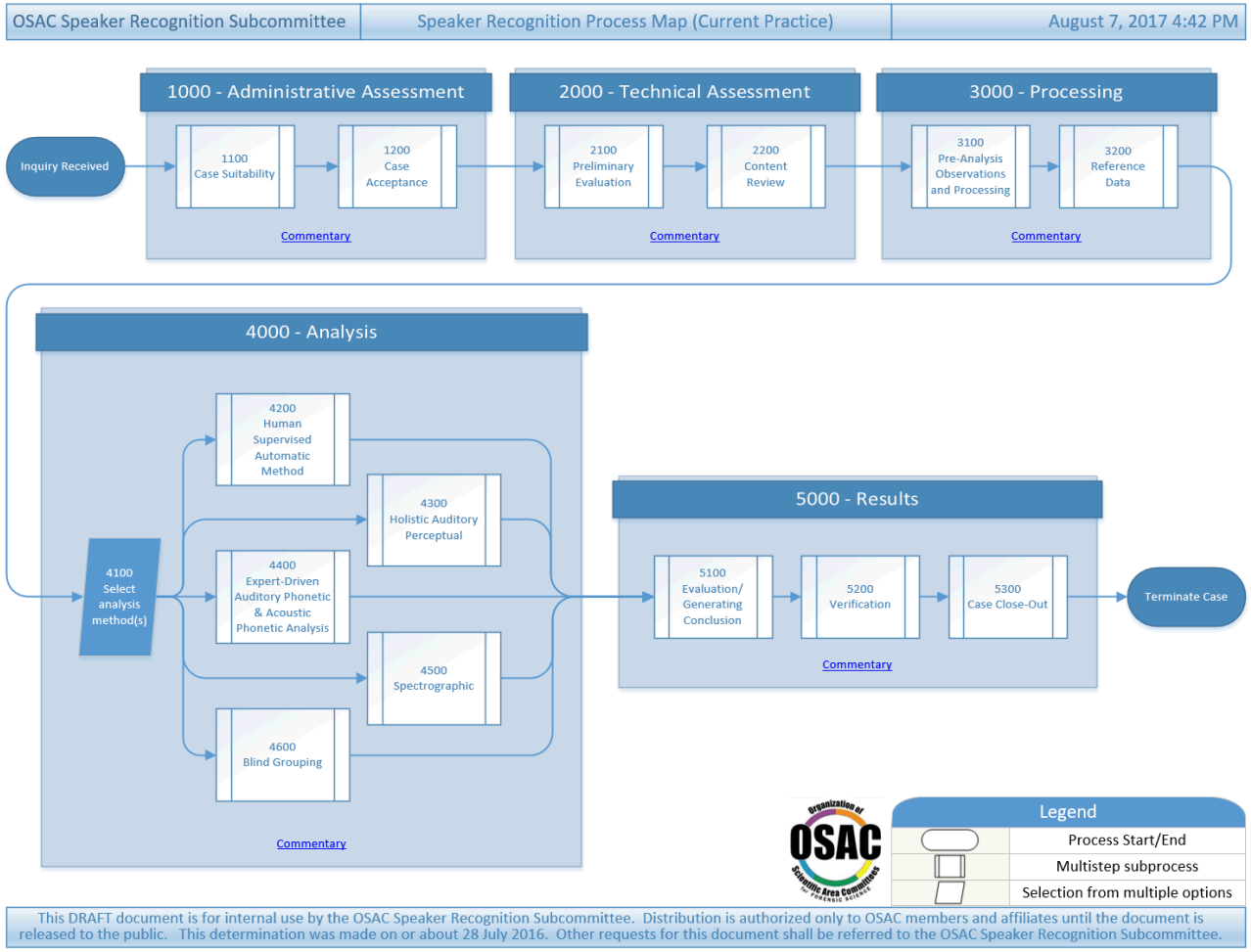
Added more methods

Noted vocabulary

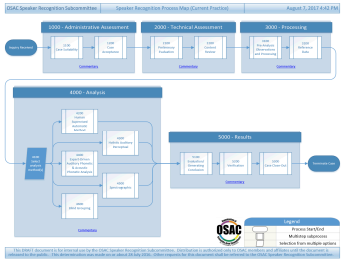
Reformatted to hierarchical structure

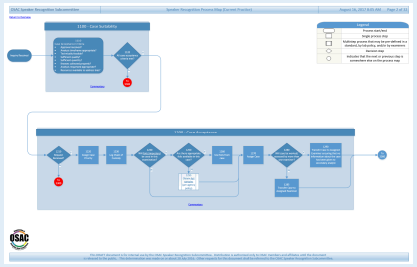
Finalized current practice (“as-is”) in May 2017

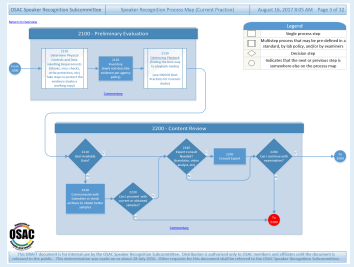
Final Process Map

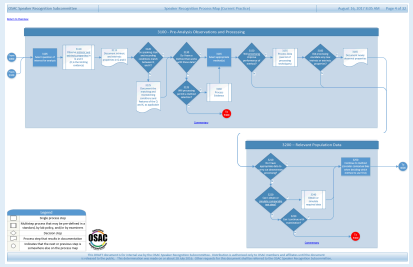


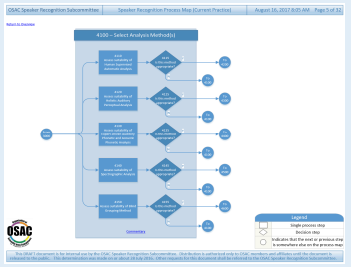
Final Process Map

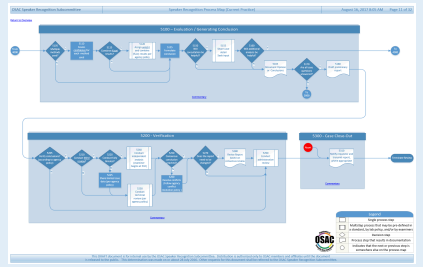


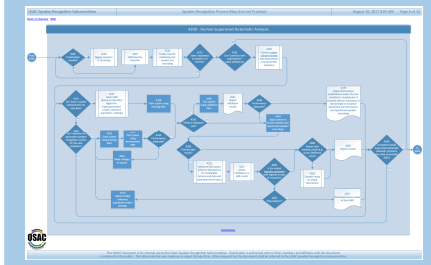


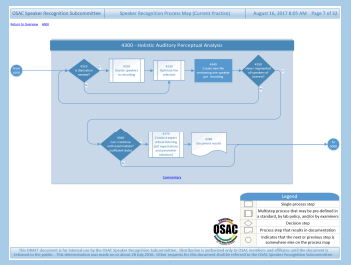


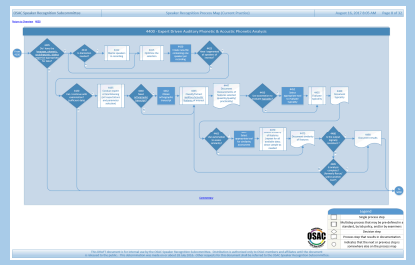


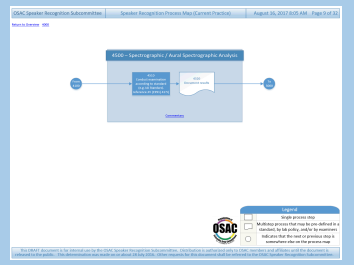


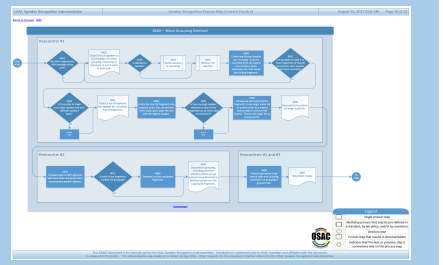












Commentary / Vocabulary

Return to Overview 3000 3100

Process Step

3100 – Pre-Analysis Observations and Processing

Description

This is the description for block 3100.

Terms and Definitions

3155: Types of processing

- Enhancement for intelligibility or listenability (e.g. tone removal, spectral shaping, adaptive filtering, etc.)
- Normalization
- Convert sampling rate / bit depth
- Channel conversion
- DC offset
- Anti-aliasing

Comments

Insert comments for block 3100 here.

Issues

Insert issues for block 3100 here. For example, is more research needed to confirm the actions described in this block?

References

Insert references and citations for block 3100 here.

Revised

August 16, 2017 8:05 AM

Return to Overview

Terms and Definitions	Terms and Definitions
Extrinsic Properties	Breathing patterns
Duration	Non-speech vocalizations (coughing, laughing)
Sampling rate	Pathologies
Stereo/mono/multi	Idiolectal terms/phrases/conversation openers/enders
Channel/bandwidth (landline, cell, microphone, transmitter, video)	
Bit depth/rate	Other Terms
File format	Human Supervised Automatic Analysis
DC offset	Holistic Auditory Perceptual Analysis
Encoding (linear, mu-law, alaw)	Expert Driven Auditory Phonetic & Acoustic Phonetic Analysis
Medium (through air, wall, window)`	Spectrographic Analysis
Codec	Blind Grouping Method
Compression (type)	
Reverberation	comparable data
Distortion	critical listening
Noise	foils/imposters
SNR (how measured?)	mark/marking
Drop-outs	optimize
Clipping	questioned/Q
Interference	known/K
Speed	match/mismatch
Proximity to microphone	processing
Overlapping talkers	simulate
Metadata (timestamp, source, etc.)	test data
Environmental noise (wind buffeting, rain, machine, media)	contextual bias
Aliasing	diarization
	segment
Intrinsic Properties	sufficient data
Language	tune
Gender	case data
Accent/Dialect	ASR
Articulation	classifier
Fluency	algorithm
Speaking/Articulation Rate	hyperparameter
Voice quality	model
High/low vocal effort	reference/relevant population
Emotional state/condition of the speaker (intoxicated, fatigued, illness, angry, medicated)	settings
	comparable system
Physical activity	error rate
Stress (cognitive, emotional, physical)	confidence
Speech type (read, repeated, oratory, conversation, voicemail, conspiratorial)	likelihood ratio
Obscured speech (mask, other face covering)	score
Voice Quality (modal, whisper, shouting, falsetto)	document results
Awareness of being recorded	document issues
Interlocutor relationship	phonetic, sociolinguistic, dialect expertise
	orthographic transcript
Features often used in Acoustic/Phonetic/Expert Listening/Statistical	auditory/acoustic features of interest
Voice onset time (acoustically measured)	classify
Pitch (F0 analysis)	typicality
Pronunciation (by listening, by acoustic measures - disfluencies)	similarity
Articulatory features (Ex: vowel analysis by 1) listening judgments, 2) acoustic measurements)	fuse
	weight
Prosodic analysis (speaking rate, articulation rate, pause rate, pause type – filled/unfilled, fluency, intonation, false starts, lengthening of syllables)	agency policy
	blind review
Voice quality (falsetto, fry, breathiness)	independent analysis
Vocabulary usage	consensus (review)
Dialectal/sociolectal analysis	administrative review
Foreign-accented	acceptance criteria

Summary

Stakeholders

- All groups

Distill key steps

- Reecognize differences
- Consensus, not necessarily100% agreement

Vocabulary

Follow-on work

- Use PM as baseline for discussion on recommended forensic processes
- Develop a “recommended practice” (i.e. “should-be”) process document
- Release as an OSAC document