Comparative Feature Extraction Techniques in Speech and Radar Emitter Recognition

Robert A. Newport
robert.newport@hdr.mq.edu.au
Supervisor: Dr Len Hamey

Research Frontiers II, MRes Yr 2

April 13, 2020

Abstract

Since the 1950s both speech and radar emitter recognition methods have independently adopted descriptors in an attempt to facilitate feature extraction and recognition. Contemporary methods for recognition both depart from the use of these descriptors. However, a lack of interdisciplinary citations show that these methods can benefit from cross-fertilization.

1 Introduction

In the 1950s, interfacing with a computer meant creating punch cards and configuring switches. Even today, the use of a keyboard and mouse requires some degree of training. However, interacting with many of the new home devices that use speech recognition shows how a speech interface can make devices easier to operate. Creating this simpler verbal interface may be one of the reasons researchers pursued speech recognition. The development of the Kay Electric Co. sonogram in 1951 allowed speech recognition to empirically measure sounds for the first time. This led to more sophisticated methods using spectrograms, which measure the intensity of frequencies over time.

In the seemingly totally different field of Electronic Warfare (EW), the timely recognition of waveforms can prevent a Radar guided surface to air missile (SAM) from reaching its target. In fields like EW, not only is waveform recognition critical for identifying radar emitters for battlefield intelligence, but also the jamming of these signals for stealth. These *cat-and-mouse* style games were primarily developed during and after America's Vietnam War and are suitably called Electronic Counter Measures (ECM) and Electronic Counter Countermeasures (ECCM) for both the avoidance of radar detection through trickery, and the avoidance of the trickery itself. Identifying the emitter of a waveform sample involves training a machine learning algorithm to classify features extracted from the samples. Each sample's spectrogram is highly complex; therefore, some generalization of the data will be required to provide features for training that do not lose their statistical relevance due to signal data being spread thin over a large number of time samples and frequency bins, resulting in a phenomena called *data sparsity*.

Interdisciplinary advances in the physiological understanding of the human acoustic process were explored to understand how consonant data derived from a spectrogram can be clustered for the purpose of feature investigation. It was theorized in the 1950s that amplitude-quantized samples captured through sensory transducers could abstract raw data into meaningful perceptual coordinates[10] — one application of which could be to serve as an interface for communication between humans and machines. In this seminal period, characterizing distinctive shapes within input samples was found to be invariant over a large number of samples and thus, signal feature extraction was born. Research quickly found that methods using statistical analysis could be influenced to self-organize, adapt, and learn to correctly classify features extracted in different ways. The convergence of these early studies provided a foundation for developing descriptors used to increase the accuracy of early computed speech recognition.

The following paper intends to step the reader through a comparative analysis of various feature extraction methods used in acoustic and Radar source recognition tasks. To do this, a background chapter will illustrate research from inception into maturity by examining how Radar and speech feature descriptors were developed to fit into software parameters in early microprocessors. Then, a chapter on contemporary methods illustrates how a departure from those traditional methods is leading to a convergence of auditory and Radar analysis through military biomimetic whale sound research, and how similar convergence can be used as a Electronic Counter Countermeasure through Random Spectral Feature (RSF) analysis.

2 Background

This chapter describes the use of descriptors by both Radar and speech recognition to facilitate feature extraction in the early phase of speech and radar recognition development. At its most basic, descriptors are names used to describe a unit of stored data. They strive to capture features across a group of samples, in a repeatable way, to be stored in a database. The descriptors became a tool used by researchers eager to adopt a common way to capture features and instead focus research on advancing classification algorithms. In the quickly accelerating radar emitter recognition field, Pulse Descriptor Words (PDW) became the standard data format for sharing known emitter information.

2.1 Sound frequencies: discovery to descriptor

In the 1930 introduction to his paper on Human Speech[27], Sir Richard Paget described sounds as "small repeated pressures – within certain limits of frequency – on the drum of our ears." His paper provided a seminal background for future work in speech recognition by providing "landmarks in vocal acoustics" through a description of the physiological connection between how humans make sound, how the sound is heard, and the actual physics behind sound as a medium.

As computers gained adoption in the 1950s, "automatic sensing" described by David[10] as an early form of speech recognition, was a task to be completed "before the Russians," and a concerted effort was made to connect the physiological research done by Paget with the current understanding about sensors and computers. Technology at the time allowed for the digitizing of sound through quantized amplitude samples output from a transducer; however, David accurately predicted that simply having more data through digitization did not necessarily move towards the goal of classification and transcription. Instead, he suggested abstracting the raw data found in representations, such as Radar signals and speech waveforms, into "meaningful perceptual coordinates." He further hypothesised that the electronic template matching systems made by engineers might be superseded in accuracy by examining features or properties distinct to a shape or waveform. By replacing patterns with features, he argues that the issue of variability is decreased, since things like speech and handwriting features are relatively more invariant than the digitized patterns themselves. Furthermore, he astutely understood that the selection of appropriate features suitable for recognition was "least mentioned" in research of the period.

As the understanding of both waveforms and machine learning matured, research moved towards organizing speech waveforms into meaningful descriptors. In the field of sound recognition, O'Shaughnessy[26] attempted to take previous research combining physiological characteristics and feature exploration[27, 10] one step further by identifying and isolating distinguishing features into clusters based on differences in the characteristics found in voicing and duration. O'Shaughnessy saw that a durational model could be applied using both an "articulatory mechanism" based on motion restrictions on the lips and tongue, as well as a "phonological mechanism" based on duration as a cue used to acoustically punctuate perception. Using this model, O'Shaughnessy created a descriptor of these "consonant clusters" by using letters to represent syllables in a word, making a cluster-vowel-cluster (CVC) descriptor to represent a basic English word[26]. For example, the word *splits* is a CVC containing /spl/ and /ts/ consonant clusters. Using this descriptor, more complex structures like VC_1C_2V double clusters could be made and compared to CVC contexts with the durational changes of C_1 and C_2 viewed as measurable effects of the insertion. Digitizing each word was done via a 7 kHz wide-band spectrogram segmented using an overlay template, with the lengths of consonants converted into time and rounded to 5 ms. Problems with dipthongs and fricatives presented challenges for the model; nevertheless, this research paved the way for further study into how sound and speech could be measured and clustered in order to extract features and classify results.

2.2 Radar frequencies: discovery to descriptor

In the late 19th Century, German scientist Heinrich Hertz[5] first demonstrated that objects made out of metal would reflect radio waves in a predictable fashion. Shortly after, in 1906, Christian Huelsmeyer patented a means of detecting ships obscured by fog using radio waves[18]. However, these methods were restricted to short ranges and low power, making them unsuitable for wartime use against high flying bombers. During World War II, British scientists used the earlier research by German scholars Hertz and Huelsmeyer to develop the method of using an oscilloscope to accurately time pulse reflections in order to estimate distance and position[34]. The name

RADAR comes from the acronym that describes its basic purpose: RAdio Detection And Ranging.

After WWII, as more countries developed their own Radar systems, the need to identify Radar signals as an instrument of Electronic Counter Measures was becoming an increasing necessity. Hertz's early experiments demonstrating that radio signals were an electromagnetic waveform opened up research into descriptors for categorizing and identifying unique signatures in the emissions. In July 1965 during America's conflict over North Vietnam, an F-4 Phantom II and a Navy RA-5C reconnaissance aircraft were both downed by Radar-guided Russian SA-2 Surface-to-Air Missiles (SAM). With the loss of intelligence and cost of the RA-5C over 5.5 million dollars, an emergency seminar was held and an operation, known as Wild Weasel[37], was awarded to the Applied Technology corporation for the development of the AN/APR-25, the first Radar Warning Receiver (RWR).

The purpose of an RWR system is to warn the pilot that a Radar has targeted the craft. The AN/APR-25 does this through on-board *crystal video receivers* on all sides of the craft, with the rudimentary assumption that the emitter source location correlates with the side containing the earliest detection. When the hostile Radar emitter's transmission impinges upon the AN/APR-25 wide band receiving antenna of the crystal video receiver, it feeds the signal through amplifiers and filters tuned to consecutive fragments of the covered band. This permits the concurrent discrimination and reception on multiple parts of the band allowing for rapid identification. However, before this happens, the various different emitter waveforms have to be parsed from one another. Doing basic pattern matching would fail due to the unintentional modulations found as artifacts from electrical power amplifiers and other components in the hardware. Instead, the AN/APR-25 measures slices of the frequency over time waveform, as a rudimentary spectrogram, and compares those to attributes to those of known enemy Radar emitters. In 1970, American patents[11, 3] illustrated the various ways Radar emitters could be identified by duration (pulse width) and frequency characteristics within each cycle (pulse repetition interval) creating a race for military powers to format and organize Radar emissions in order to be able to store them in databases for real time software enabled recognition. The increasing penetration of computers and software systems in the late 1960s and early 1970s produced a capability to quickly compute probabilities for matches from a database of characteristics, if those features are organized uniformly for database storage. In 1972 the ALR-45, based on a *CPU and software*, was able to execute probability statements[37] based on Pulse Repetition Interval, Pulse Coding, Frequency, Pulse Width, and other emitter-specific attributes. This system worked exceptionally well, and lead to a 1976 patent[23] explicitly outlining the importance of "digital words" to simplify the processing of emitter recognition by organizing various Radar attributes into comparable waveform characteristics. As the pursuit of Radar emitter recognition advanced, the digital words described in this patent become known as Pulse Descriptor Words in the Electronic Warfare research field.

3 Traditional features

Descriptors used in both speech and Radar recognition can be seen as a product of the technology used to exploit them. Increased stored descriptor data of consonants and radar emitters made accurate automatic recognition an increasing possibility if clusters of those features could be compared in a statistically manner on a computer. In the previous chapter, the jump into *CPU and software* based Radar devices such as the ALR-45 were a bridge that compelled Applied Technology to cross from a circuit design company to a computer sciences company[37]. The following two descriptor formats describe how both Radar and speech could fit into the emerging software framework by leveraging their respective data formats.

3.1 Automatic Language Identification (ALI)

Vibrations caused by human biology to produce sound for use in language is a particularly challenging model to capture due to the variety of features that could be used in a model. Furthermore, some languages may have features that are not shared by all, producing gaps in the reliability of models when trained. For example, Japanese speech uses a *sentence-final particle* instead of an interrogatory inflection when asking a question. In English, the statement "Cup of tea?" with a change in tone at the end of the sentence can imply that the speaker is asking if the target is interested in having a cup of tea. In spoken Japanese, the same intent is conveyed using a *-ka* word stem at the end of a sentence with no auditory inflection[2].

In order to research these intricacies, in the 1980s Li et al[22] segmented speech into its acoustic phonetic counterparts with the goal of creating a statistical model for use in automatic identification of the language being spoken. Six acoustic-phonetic classes used to distinguish between five different languages in the study included Syllabic Nuclei, Non-vowel Sonorants, Vocal Murmur, Voiced Frication, Voiceless Frication, and Silence and Low Energy Segments. Statistics of each recognized language's segments were used to train finite state models. These models were used on strings (as syllables) and individual concatenated segments. Syllabic boundaries were not used. Instead, the models described statistics for either inter or intra syllable nucleus segments, including their durations.

Before analysis could begin, preprocessing would extract features such as power, voicing, and zero-crossing count in order to brings samples to a comparative state. After preprocessing, the five step segmentation would begin. Firstly, a frame rate of 10 ms to classify silence, voicing, and frication was created. This demarcated a constant area of the signal to be analyzed. After the frames were set, they were encoded into one of the classes shown in the list above. This did not classify the segment yet, since segments sometimes contained multiple concatenated frames. After scanning multiple frames, the ones that shared similar classifications were formed into a segment. Very short segments made from the frames concatenated in the previous step were time smoothed in order to eliminate transitional or extra segments. Then finally, vocalic segments were broken up using non-vowel and syllable nuclei sonorant demarcations. This led to an accuracy of 80 percent across 50 people talking in the research that first illustrated this technique[22].

3.2 Pulse Descriptor Words (PDW)

In the same way that buildings, mountains, tanks, and aircraft carriers reflect light, they also reflect radio waves. Like light, a portion of radio waves emitted from a Radar transmitter are bounced back[36]. At its most rudimentary, a Radar receiver can calculate the length of time it takes for a round trip between transmission and reception to determine the distance a target is from the device. Different factors affect attributes of the Radar signal, including its range, resolution, and strength. These factors could include transmission power, transmission duration, antenna size, number of search scans of the area, wavelength, and different types of pulse compression[36]. Traditional methods used in *a priori* database filtering for identifying Radar signals include the following five classical parameters[25]: Pulse Width (PW), Direction of Arrival (DOA), Time of Arrival (TOA), Pulse Amplitude (PA), and Radio Frequency (RF). These parameters in Electronic Warfare systems operate at 500k to 1M pulses per second[13] making operations complex to parse. A traditional format used to describe an emitter is the Pulse Descriptor Word (PDW), which may contain some or all of these parameters[35]. Each detected pulse measurement can include the bearing, center frequency, time of arrival, pulse width, and amplitude converted to a format compatible for digital processing as features.

Traditionally, these signals are packed into a PDW, an Electronic Warfare receiver sends them to a pulse-sort processor for Pulse Repetition Interval (PRI) identification and modulation isolation. They are then compared to an existing emitter database[6]. Random modulations like jitter, stagger, and switching aberrations may interfere with the correlation of emitter data based on frequency, pulse width, and PRI causing final identification to generate an emitter list with a variable recognition score. To improve these scores for emitter recognition, researchers in the field use machine learning to increase the efficiency and accuracy of the Electronic Warfare recognition process.

4 Beyond Traditional Features

As described in the section above, Pulse Descriptor Words were a traditional format used in *a priori* filtering and classification of Radar emitters during the middle formative stage of Radar recognition research. Speech and language research also focused on fixed descriptors in order to capture single word voice, multi word voices, or spontaneous voices[1, 20] all using a fixed set of descriptors like the ALI paper referenced in the previous chapter[22].

Both techniques created demarcations, like the 10 ms ALI frame and the Radar defined pulse width, and then defined each segment in order to facilitate recognition of the overall signal. Unintentional jitter in Radar signals, like transitional segments in short ALI signals, were smoothed out and discarded.

However, with the current densely populated battlefield of agile Radar emitters able to stagger and jitter pulse repetition intervals on a pulse by pulse basis, as well as the natural differences in the way humans speak and vocalize, there is increasing pressure to break out of the rigid descriptors and into more inclusive types of features for more granular recognition. In Radar, this includes the ability to detect specific emitters within a type, similar to speaker recognition withing the field of speech. This opens up new areas of application in both fields, including speaker authentication and forensics[28], biomimetic sonar[38], and mobile emitter tracking[14].

This section will illustrate shared techniques and methods used to isolate, segment, define, and extract features used in both speech and Radar recognition. Guo et al[14] stated that feature extraction in Radar recognition can be generalized into the following three categories which we suggest can also be applied to speech recognition:

1. Classic parameters (mentioned in previous section) matched to waveform data obtained through prior knowledge.
2. Raw signals as inputs for deep learning.
3. Intrapulse signals for feature analysis.

In Radar recognition, a break from the reliance on the five traditional parameters by many of the contemporary research papers included in this study implies that research into individual Radar features is becoming increasingly popular, especially since many traditional methods can barely meet the requirements of warfare with signal-to-noise performance not greater than than 10 dB[16]. An even further departure is seen within the field of speech recognition, where the source no longer has the constraint of being human language or biologically created. Animal recognition through vocalizations, gunshot recognition, and emotion detection spur the progress of early speech recognition into new areas with increasing requirements to be investigated in the next subsections.

4.1 Unintended and Spurious Signals

Unintended Modulations on Pulse (UMOP) within the scope of Radar signals include the unwanted and unavoidable yet persistent waveform features found in an emitter signals that are idiosyncratic to that particular emitter's transmitter. This devices may include nonideal and nonlinear electronics like the oscillator, power source, or especially its power amplifier[14]. UMOP characteristics can include frequency drift, pulse envelope, and phase noise.

Drift from the nominal frequency can be caused when the emitter changes frequency, as well by aging and temperature changes in the hardware. Pulse envelope changes are inevitable artifacts from unintended and unwanted attributes such as lead inductance and capacitance distribution from varying current and working voltage. Phase noise is a ratio of signal to noise power measured in the frequency domain. These unintended modulations and perturbations can be scrutinized in other research areas to yield similarly interesting features used to recognize unique attributes belonging to the signal's source.

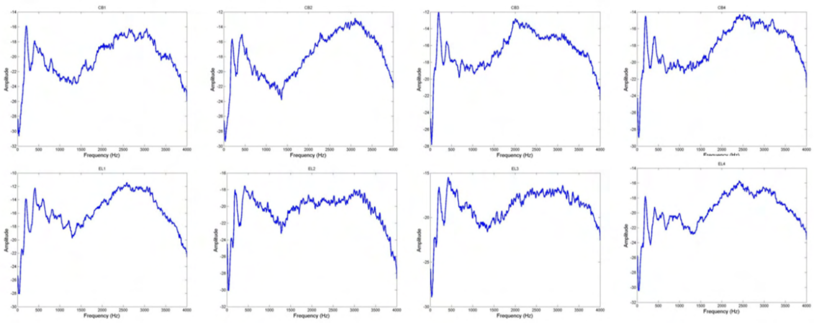
Within speech recognition, unintentional features including jitter, perturbation quotient (PPQ), harmonics-to-noise ratio, pitch, and shimmer[7, 29, 30, 24] originating from the vocal source. These features could be caused by heightened emotions, speech impediments, accents, and other types of dysphonia. When clinically assessing the quality of a patient's voice, descriptive parameters are required to correlate the pathological voice with acoustic phenomena. One such study[24] attempted to correlate voice abnormalities with airflow anomalies over the glottis caused by the morphology of particular vocal folds. The significance of this is that the organic pathology of the dysphonia stems from the vocal tract configuration rather than the usually assumed vocal source. This research means that not only are discoveries made in the automatic detection of speech dysphonia, but they are also made in the organic pathological sources for the dysphonia itself.

A military bomber must avoid radar guided missiles while using its own radar to identify and neutralize potential threats. If a radar guided missile is deployed, jamming or trickery must be used to avoid being shot down. This may include identifying the radar emitter used and returning a jamming frequency to prevent an accurate location from being returned. Being able to recognize the type of emitter must move beyond simply identifying the PDW since many different Radar emitters may emit the same frequency. By understanding the fine differences between emitters that produce the same frequency, the specific hardware producing the frequency can be known, allowing for additional intelligence on the battlefield. For example, a bomber avoiding a Radar guided air to air AIM-7E2 missile will know it has a to 2 second warm up time based on radar emitter recognition intelligence. This intelligence is critical to mission success in warfare and the scrutiny of the unintended features used to identify emitters has created a sub-field in Radar recognition known as Specific Emitter Identification (SEI) which is increasing in importance in its role in Electronic Warfare. Especially where exotic emitters create chaotic battlefield conditions where traditional techniques for isolating parameters are purposefully hidden during countermeasure operations[21]. Early examples of using SEI have achieved high levels of accuracy, including research in a flight test program using "the fine frequency structure at the pulse leading edge" achieving 90%-95% confidence levels.[21, 32]. Some research into alternative techniques for extracting UMOP have yielded positive results, especially with the Ambiguity Function (AF) introduced by Gillespie and Atlas[12] to Radar emitter recognition. Guo et al[14] were able to use AF to isolate UMOP features by enlarging differences using a Discrete Wavelet Packet Transform (DWPT). This involved a three step process: the signals are first mapped onto a time delay and frequency offset plane; an appropriate diagonal slice extraction from the Ambiguity Plot is found with the most information in the time to remain space; since the diagonal slices are all very similar, the differences are enlarged using a Discrete Wavelet Packet Transform (DWPT). It is these differences that aid in the feature extraction using this method.

A similar approach is seen in the new field of digital speech forensics. As a response to the abundance of tools accessible to malicious amateur and expert forgers, this field attempts to extract forensic evidence embedded in the signal processing chain of recording and amplification devices in order to recognize discrepancies, edits, or sources of audio samples[28]. Microphones, amplifiers, and other electronics in the audio capture, recording, and broadcasting chain are susceptible to the same spurious modulations that provide UMOP the opportunity to profile specific hardware. By examining Labeled Spectral Features(LSF) and Random Spectral Features(RSFs) of the audo sample, distinguishing characteristics can be used to build a dactylogram of the voice recording and broadcasting chain that could be used to identify the hardware. Figure 1 illustrates some unique visible characteristics over eight different handsets with the same speech utterance recording. The study used speech recordings from 8 Plain Old Telephone Service (POTS) devices 53 male and female speakers. SVM reported the greatest accuracy in device recognition at 97.58%. However, studies using a greater number of POTS devices will be necessary to see how this model will scale will a larger number of profiles.

In Radar UMOP, only one research study was found to test UMOP characteristics using a high (411) number of labelled hardware emitters for the purpose of verifying the validity of UMOP proposed features[32]. The study found that 0.2 μs of the leading and trailing edges provided the most stable and consistent for UMOP inspection, with unintentional modulation on frequency more obvious than amplitude[32].

Figure 1: Eight different phone handsets, mean spectrogram[28]



However, due to these features being dependent of real world hardware profiles and characteristics, it will be difficult to sufficiently investigate the effectiveness of this technique without a large amount of testing on actual devices, which is lacking, at this current phase in research. Additionally, some studies do not include the number of emitters using in testing[21, 32]. Even though these findings illustrate a great deal of promise in future work, with one such study having 104 emitters recognized with 92% accuracy[32], many new papers using large numbers of hardware emitters for testing remain to be seen and may be practically prohibitive in civilian arenas.

4.2 Acoustic and Radar Convergence via AF Analysis

If an echo is simply the reflection of a signal, the returned signal will simply be a copy. However, the delay related to the object's distance τ and the shift in frequency due to object velocity f will have an effect on real world signals coming back. The Ambiguity Function is used to scrutinize returned signals that sometimes look and behave like other signals when they are returned. A classical definition of the Ambiguity Function is defined as a 2D function of time delay versus Doppler frequency to highlight pulse distortions based on a matched filter in a receiver[39]. This is due to the pulse from the emitter changing its time delay based on distance (τ) and its frequency based on velocity (f) when it reflects off an object. As an example, if the original waveform is $s(t)$ the signal (not including noise or other contamination) reflected would be:

$$s_{\tau,f}(t) \equiv s(t-\tau)e^{i2\pi ft}$$

Doppler distortions found in frequency changes and time delays are compared to properties of the matched filter and pulse. A match is positive if signal $s_{\tau,f}$ correlates strongly with Doppler and delay (τ, f). Due to calculations sometimes producing false positive results, and due to those results being highly correlated to delays and Doppler shifts in the calculations for specific signals, the results are deemed *ambiguous*, hence the name. Each waveform $s(t)$ has a different ambiguity function, since each waveform contains its own set of false positive matches.

Even though AF was first discovered to be a novel approach to sorting Radar emitter signals almost a decade ago[17], the use of AF on intrapulse Radar and speech data still sees novel applications in contemporary applications. In one such Radar study[17], an AF structure was sliced into a 2D graph and processed using a wavelet transform and sorted. This produced high accuracy above 90% but was highly computationally expensive[17] making it prohibitive for real time processing. This is problematic in real world EW scenarios but not so much in animal speech recognition.

Bird calls, just like human speech, encode a complex framework of biological systems and auditory signals that make up a collective repertoire of songs and syllables. Spectrogram cross-correlation (SPCC) or frequency bandwidth and time duration can characterize uncomplicated song patterns from different bird species with sufficient results. However, *within-species* automatic clustering of syllables for the purpose of calculating size of the repertoire remains a challenge. As an alternative to smoothed distribution spectrograms, one study proposed the use of the Wigner-Ville distribution for feature extraction[33]. Wigner-Ville can act as a type of transform in time-frequency analysis, producing higher clarity and preventing leakage more than Short Time Fourier transforms (STFT). Evaluation of the method was applied to three hand sorted syllable classes recorded from one source. The sample size may be relatively small compared to similar studies used in Radar emitter recognition but the True Positive Rate (TPR) using the proposed method was highest among other methods tested in the study[33].

The high correlation of AF structures for electromagnetic waveforms of a specific shape can also be exploited as features for identifying Radar emitters. The *main ridge slice* from a resulting AF 3D structure using selected rotation angles and cloud structures was generalized into feature vectors for their classification and recognition using machine learning algorithms such as Kernel Fuzzy C-means in one such study[15]. The *main ridge slice* in this case is defined as where the ambiguity energy is found in the main distribution area. Intrapulse features are an important aspect of AF's ability to isolate features, and can also be used to test the velocity and range resolution in active sonar. When signal to noise is low, operations on the *main ridge slice* of an AF structure can reduce noise in novel ways. Singular Value Decomposition (SVD) can be used to denoise the *main ridge slice* before selecting the symmetric Holder coefficients and rotation angle[16]. Denoising operations using SVD are particularly effective due to different parameters and modulation types having ambiguity characteristics unique to their representative waveform[16]. Out of six signals tested using SVD over the *main ridge slice* using an AF, accuracy is seen between 88% and 100% at 0 dB SNR. At 4 dB SNR 100% accuracy is achieved across all six waveform types tested including CW, LFM, BPSK, QPSK, M-SEQ, and BFSK[16]. However, the complexity of using this method is high, particularly for short signal lengths.

As seen in the previous section, studies in bird vocalizations have borrowed heavily from physiological research into human speech biology. One step further is the understanding of biological vocalizations of marine mammals, and how those vocalizations may assist them in hunting, mating, and communication. An interesting intersection in organic vocalization and sonar research is in the military adoption of whale frequencies for the purpose of concealment and attenuation mitigation. Much like Radar, active under water sonar systems send (acoustic) signals in order to analyze the backscatter information. In a similar fashion, whales can locate prey as small as one meter from a distance of several hundred meters using click vocalizations with lower frequency vocalizations measured up to 300 km. One such study[38] found that super low frequency blue whale mating calls can travel long distances with little attenuation compared to modern active sonar[38]. Furthermore, since whale vocalizations are natural marine mammal sounds, biomimetic military applications using these sounds in an active sonar application have the added benefit of *hiding in plain sight*. Humpback whale vocalizations are a combination of complex frequency modulated (FM) and constant frequency (CF) elements between 8 Hz and 8 kHz[4]. This range is similar to the frequency band used by low frequency active sonar and has inspired a sonar model based on whale acoustic properties. In order to analyze the biomimetic active sonar whale model, a calculation is made of the transmission loss and the sound unit's ambiguity function. A rational design waveform is critical for reducing the influence of reverberation caused through the ocean by volume scatter through the body or surface scatter off the top or bottom, which is especially severe in shallow regions. The ambiguity function's echo output using an ideal channel matched filter shows similar parallels to the anti-reverberation abilities of the whale's waveform signal[38]. Therefore the ambiguity function of the calculated signal from the whale can test resolution over range and velocity by looking for either an increase in the time delay or frequency shift, or a high decrease of the ambiguity function. Biomimetic research such as this illustrates the interesting synergy we are witnessing in contemporary research as the fields of marine biology, biological vocalization, speech recognition, and signal processing converge.

5 Discussion and Conclusion

The synergy seen in biomimetic research using whale sounds and aided by tools typically seen in Radar analysis such as the ambiguity function strongly indicate that there can be much cross-fertilization between biological acoustic and Radar waveform research. Popular scholarly search engines indicate that contemporary (< 5 years) biologically inspired models in Radar are either lacking or confidential.

Future studies in Radar could follow the steps outlined in the preciously included papers by first finding syllabic repertoires (of bats, perhaps) through spectrogram cross-correlation in vocalizations as shown in the bird study, and then testing the effects using the ambiguity function as seen in the whale study. However, even though the whale (and other biomimetic) sounds can *hide in plain sight*, the method shown in the study on speaker recognition forensics illustrates how Labeled Spectral Features(LSF) and Random Spectral Features(RSFs) can be used to extract the unintended modulations seen in digital recordings and amplification in order to differentiate biological acoustics from digital ones.

In contemporary Radar research, experiments are primarily focused on a small set of modulation modes in low noise settings. Focusing on other modulation modes like FDK/PSK[40] and achieving higher accuracy below 10 dB would be a natural next step, since most Radar Emitter Recognition methods cannot achieve adequate warfare level accuracy below 10 dB SNR[16]. In cases where noise is introduced into samples, many tests only include Gaussian white noise[41], which may imply that the methods contained within the experiment are not suitable for other types of noise.

Current research in UMOP appears to focus on basic features within the waveform, neglecting combined or unique features[31] weighted within specific waveform types. Compared to the classical five parameters for a piori waveform table prediction, UMOP analysis has shown that features are not a priori predictable, and that the more extracted features are available, the better the performance will be in SEI recognition[8, 9, 19].

In both acoustic and Radar fields, new research could benefit from greater focus on accuracy in high noise environments (below 5 dB SNR). The *main ridge slice* transform using the Ambiguity Function allows for a large number of different noise reduction techniques as shown in previous sections and shows a great deal of promise to be used in both fields.

We have shown that the parallel progression of radar and speech recognition fields have similar technical requirements and challenges. This offers an opportunity for a higher degree of interdisciplinary cooperation by borrowing methods from areas such as unintended modulation research and waveform analysis as shown by UMOP and AF. As spectrograms become a more ubiquitous data format in both fields, the role of advanced complex frequency analysis algorithms will rise in importance which can move these seemingly disparate fields of speech and radar emitter recognition forward through shared discoveries.

References

[1] A pattern recognition model of voice-based personal verification systems for forensic applications. *IEEE Transactions on Systems, Man, and Cybernetics*, 10(9):585–588, Sept 1980.

[2] *The Cambridge Handbook of Japanese Linguistics*. Cambridge Handbooks in Language and Linguistics. Cambridge University Press, 2018.

[3] Steven A Wicks Andrew J Trochanowski. Automatic radar detection device., 10 1970.

[4] Whitlow W. L. Au, Adam A. Pack, Marc O. Lammers, Louis M. Herman, Mark H. Deakos, and Kim Andrews. Acoustic properties of humpback whale songs. *The Journal of the Acoustical Society of America*, 120(2):1103–1110, 2006.

[5] J. Blanchard. Hertz, the discoverer of electric waves. *The Bell System Technical Journal*, 17(3):326–337, July 1938.

[6] Yee Ming Chen, Chih-Min Lin, and Chi-Shun Hsueh. Emitter identification of electronic intelligence system using type-2 fuzzy classifier. *Systems Science & Control Engineering*, 2(1):389–397, 2014.

[7] D. G. Childers and K. S. Bae. Detection of laryngeal function using speech and electroglottographic data. *IEEE Transactions on Biomedical Engineering*, 39(1):19–25, Jan 1992.

[8] S. D'Agostino. Specific emitter identification based on amplitude features. In *2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pages 350–354, Oct 2015.

[9] S. D'Agostino, G. Foglia, and D. Pistoia. Specific emitter identification: Analysis on real radar signal data. In *2009 European Radar Conference (EuRAD)*, pages 242–245, Sept 2009.

[10] E. E. David and O. G. Selfridge. Eyes and ears for computers. *Proceedings of the IRE*, 50(5):1093–1101, May 1962.

[11] Denman R Elliott Gerald O Miller. Radar detection device., 07 1970.

[12] B. W. Gillespie and L. E. Atlas. Optimizing time-frequency kernels for classification. *IEEE Transactions on Signal Processing*, 49(3):485–496, Mar 2001.

[13] P. M. Grant and J. H. Collins. Introduction to electronic warfare. *Communications, Radar and Signal Processing, IEE Proceedings F*, 129(3):113–132, June 1982.

[14] H. Guo, X. Zhang, L. Yang, and S. Zhang. Improved fisher linear discriminant analysis for feature extraction of unintentional modulation on pulse by combining ambiguity function with wavelet transform. In *IET International Radar Conference 2015*, pages 1–4, Oct 2015.

[15] Q. Guo, P. Nan, and J. Wan. Radar signal recognition based on ambiguity function features and cloud model similarity. In *2016 8th International Conference on Ultrawideband and Ultrashort Impulse Signals (UWBUSIS)*, pages 128–134, Sept 2016.

[16] Q. Guo, P. Nan, X. Zhang, Y. Zhao, and J. Wan. Recognition of radar emitter signals based on svd and af main ridge slice. *Journal of Communications and Networks*, 17(5):491–498, Oct 2015.

[17] J. Han, M. h. He, Y. q. Zhu, and B. g. Zhu. A novel method for sorting radar radiating-source signal based on ambiguity function. In *2009 International Conference on Networks Security, Wireless Communications and Trusted Computing*, volume 2, pages 820–823, April 2009.

[18] Christian Huelsmeyer. Wireless transmitting and receiving mechanism for electric waves., 03 1904.

[19] Kenneth I Talbot, Paul R Duley, and Martin H Hyatt. Specific emitter identification and verification. 01 2003.

[20] P. Kumar, M. Deshmukh, and A. Kumar. A novel pitch based voice recognition model (pvrm). In *2018 4th International Conference on Recent Advances in Information Technology (RAIT)*, pages 1–4, March 2018.

[21] L. E. Langley. Specific emitter identification (sei) and classical parameter fusion technology. In *WESCON/'93. Conference Record,*, pages 377–381, Sep 1993.

[22] K. Li and T. Edwards. Statistical models for automatic language identification. In *ICASSP '80. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 884–887, Apr 1980.

[23] Oscar Lowenschuss. Electronic countermeasure system., 07 1976.

[24] M. Markaki, Y. Stylianou, J. D. Arias-LondoÃśo, and J. I. Godino-Llorente. Dysphonia detection based on modulation spectral features and cepstral coefficients. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5162–5165, March 2010.

[25] D. J. Milojevic and B. M. Popovic. Improved algorithm for the deinterleaving of radar pulses. *IEE Proceedings F - Radar and Signal Processing*, 139(1):98–104, Feb 1992.

[26] D. O'Shaughnessy. Consonant durations in clusters. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 22(4):282–295, Aug 1974.

[27] R. Paget. *Human Speech: Some Observations, Experiments, and Conclusions as to the Nature, Origin, Purpose and Possible Improvement of Human Speech*. Number v. 12 in Cognitive psychology]. Routledge, 1999.

[28] Y. Panagakis and C. Kotropoulos. Telephone handset identification by feature selection and sparse representations. In *2012 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 73–78, Dec 2012.

[29] Vijay Parsa and Donald G. Jamieson. Identification of pathological voices using glottal noise measures. *Journal of speech, language, and hearing research : JSLHR*, 43 2:469–85, 2000.

[30] Robert A. Prosek, Allen A. Montgomery, Brian E. Walden, and David B. Hawkins. An evaluation of residue features as correlates of voice disorders. *Journal of Communication Disorders*, 20(2):105 – 117, 1987.

[31] X. Ru, C. Gao, Z. Liu, Z. Huang, and W. Jiang. Emitter identification based on the structure of unintentional modulation. In *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 998–1002, Oct 2016.

[32] X. Ru, H. Ye, Z. Liu, Z. Huang, F. Wang, and W. Jiang. An experimental study on secondary radar transponder umop characteristics. In *2016 European Radar Conference (EuRAD)*, pages 314–317, Oct 2016.

[33] M. Sandsten and J. Brynolfsson. Classification of bird song syllables using wigner-ville ambiguity function cross-terms. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 1739–1743, Aug 2017.

[34] K. Schaffel and United States. Air Force. Office of Air Force History. *The Emerging Shield: The Air Force and the Evolution of Continental Air Defense, 1945-1960.* General histories. Office of Air Force History, United States Air Force, 1991.

[35] M I Skolnik. *Radar Handbook, Third Edition.* Electronics electrical engineering. McGraw-Hill Education, 2008.

[36] G.W. Stimson. *Introduction to Airborne Radar.* Aerospace & Radar Systems. SciTech Pub., 1998.

[37] Northup Grumman Electronic Systems. The radar warning story. *Asian Defence Journal*, V. 72, 1982.

[38] Q. Wang, L. Wang, and L. Zou. Whale-inspired sonar in covert detection. In *2016 IEEE/OES China Ocean Acoustics (COA)*, pages 1–4, Jan 2016.

[39] P.M. WOODWARD. 7 - the transmitted radar signal. In P.M. WOODWARD, editor, *Probability and Information Theory with Applications to Radar (Second Edition)*, International Series of Monographs on Electronics and Instrumentation, pages 115 – 125. Pergamon, second edition edition, 1953.

[40] Shuhua Xu, Lina Xu, Zhengguang Xu, and Benxiong Huang. Individual radio transmitter identification based on spurious modulation characteristic of signal envelop. pages 1 – 5, 12 2008.

[41] X. Zhang, P. Luo, and X. Hu. A hybrid method for classification and identification of emitter signals. In *2017 4th International Conference on Systems and Informatics (ICSAI)*, pages 1060–1065, Nov 2017.