Vol. 3 No.2 2025

RECOGNIZED IN "Y" CATEGORY BY

Research Consortium Archive

P(ISSN) : 3007-0031

E(ISSN) : 3007-004X

https://rc-archive.com/index.php/Journal/about

ENGLISH SPEECH SOUNDS: FOURIER ANALYSIS AND SPECTROGRAM STUDIES

Ambreen Zehra Rizvi
Assistant Professor, Faculty of Engineering, Science & Technology, Hamdard University Main Campus, Karachi, Pakistan. Email: ambreenzehrarizvi@hamdard.edu.pk

Nazra Zahid Shaikh
Senior Lecturer, Department of English, Faculty of Social Sciences and Humanities, Hamdard University Main Campus, Karachi, Pakistan. Email: nazra.zahid@hamdard.edu.pk

Publisher : EDUCATION GENIUS SOLUTIONS
Review Type: Double Blind Peer Review

ABSTRACT

In this paper, the Fourier analysis and spectrogram techniques for the study of the English speech sounds is discussed. One of the subjects that we will study in this book is the structure of some mathematical functions (namely, the Fourier transform) and some signal processing algorithms reminiscent of the inverse Fourier transform. The paper shows the significance of spectrograms as key visualized representations of acoustic phonetics with which to visually inspect formant structures, voice onset time, and other key temporal-spectral attributes of English phonemes. We use our techniques to demonstrate experimentally the acoustic signatures of vowels and consonants. The paper also covers potential applications in speech recognition, linguistic studies and speech pathology diagnosis. What we find is that, even in an era of increasingly sophisticated acoustic analysis, the Fourier-based spectrograph plays a vital role within modern phonetic research by detailing perceptual principles which cannot be accounted for on the basis of audition alone.

Index Terms: Fourier transform, spectrogram, acoustic phonetics, speech signal processing, formant analysis.

Introduction

The study of speech sounds of a scientific nature has changed dramatically as a result of the introduction of sophisticated mathematical methods of signal processing, of which Fourier analysis has been particularly wellassembled to be sure pm 854 accepted. First formulated in the early 19th century by Jean-Baptiste Joseph Fourier, this mathematical system has been applied with great success in the domain of acoustic phonetics, providing precise and quantitative means of decomposing and representing the complex waveforms produced by human vocal tracts. The capacity to analyse and extract components of complex acoustic signals into the elementary frequency components have given an unsurpassed view in the process of speech production and perception.

In the present day, phonetic research makes extensive use of Fourier-based analysis of the spectral properties of speech sounds. These techniques allow the systemized investigation of several sound parameters, which contribute to phonemic differentiation in English and other languages. By transforming temporal waveforms into spectral representations, phonetic properties can be objectively measured and compared that would be impossible to perceive through auditory observation alone. This algebraic machinery has thus played a fundamental role in both theoretical linguistics as well as in applied speech sciences.

The analysis and application of Fourier analytical methods to spectrographic analysis of English phonemes are examined in detail in the present study. After a review of the Fourier transforms, with a presentation of the underlying mathematics, the issue moves to the quantitative methods for exploring it digitally in the context of recent speech systems. The methods for the time-scale and time-frequency approximations that are in use for non-stationary signals, specially for speech signals is addressed, with focus on the computational requirements which have to be met for efficient analysis.

The main body of this text is concerned with the description of spectrographic displays of several categories of English speech sounds. The detailed acoustic characteristics of vowels, consonants, and the prosodic features are studied by the experimental speech data analysis. The review emphasizes the ways in which spectrographic visualization can uncover important phonetic information, including formant structures, voice onset properties, and temporal-spectral patterns that differentiate phonemic categories.

In addition to the basic phonetic analysis, this article also discusses in general the potential of Fourier-based techniques for phonetic research and speech technology. Applications in dialectology, language acquisition studies and clinical phonetics) and adoptions in automatic speech recognition and speech synthesis systems are also discussed. These applications illustrate that Fourier methods are still important tools in both the academic and industrial speech science communities.

The combination of mathematical signal processing and linguistic analysis is a major step forward in the science of speech sounds. Fourier analysis was used to develop objective, quantitative accounts of speech production, leading to the more articulate description and comparison among languages of lexical items (e.g., vowel spaces). Such methodology has proven to be of especial importance in experimental phonetics, where it aids in the critical testing of hypotheses and the rigorous formulation of theory.

Mathematical foundations of speech and language Prosody Basics of the Fourier transform

Fourier transform provides the basic mathematical tools necessary for transforming time-domain waveforms into frequency-domain waveforms. As we will see, this transformation is particularly useful in speech analysis, since it lets us reduce any complex acoustic signal to simpler sinusoidal components. The continuous Fourier transform of an analog speech signal x(t) may be denoted mathematically by the integral equation:

$$X(f) = \int[-\infty \text{ to } \infty]\ x(t)e\text{^}(-j2\pi ft)\ dt$$

This equation is a discrete description of the time-dependent signal transformation from time domain to frequency domain, where –foreach frequency f (represented by X(f)), three major parameters are given: frequency, amplitude and phase. The complex-valued function X(f) fully retains the information of the original signal with respect to its frequency content in the entire frequency range.

In Digital Signal Processing, where speech is sampled at equally spaced intervals of time the DFT is the realisation of the above concept. The DFT equation associated with a digital speech signal x[n] (having N samples) is then described by:

$$X[k] = \Sigma[n=0 \text{ to } N-1]\ x[n]e\text{^}(-j2\pi kn/N)$$

where k is the discrete frequency index. This transform consists into mapping the N sample time- domain sequence in a N sample frequency-domain sequence, where each X[k] represents the values of one frequency binlocation k · fs/N; where fs represents the sampling frequency.

The high computational complexity of the DFT was later addressed in the Fast Fourier Transform (FFT) algorithm, that reduced the number of operations from $O(N^2)$ to $O(N \log N)$. This effect has led to real-time spectral analysis of speech signals possible using limited desktop machine tools. The FFT algorithm does this by taking advantage of symmetry and periodicity properties of the basis functions, and breaking the computation of the DFT into smallers problems in a recursive manner.

There are a number of important issues when using Fourier methods to speech signals. The resolution in frequency depends here solely on the length of the window; larger windows offer higher resolution in frequency, however, they might also hide rapid changes over time. Conversely, shorter windows provide better time resolution but worse frequency resolution, which captures the essence of the time-frequency uncertainty principle in signal analysis.

The window functions used in the spectral analysis actually matter. Classical windows like the Hamming, Hanning, and the Blackman window are used to reduce the spectral leakage that result when the observation interval is finite. These window functions smooth the signal at the boundaries of the analysis window, limiting spectral leakage and artifacts in the frequency domain, while imparting their own characteristic spectral footprints that need to be considered when interpreting results.

Due to the mathematical behavior of the Fourier transform, it is particularly well suited for examining the quasi-periodic components of voiced speech sounds, and the harmonic characteristics of the glottal excitation being reflected in the frequency domain. On the other hand, for voiceless speech sounds, the Fourier transform gives information on spectral-envelope and on energy distribution in the frequency domain as it does for voiced ones but the interpretation of the results needs some different considerations.

Short-Time Fourier Transform (STFT) for Speech Analysis

The speech signals have non-stationary nature with fast changing spectral information of sounds with the passage of time. Consequent Fourier analysis is not adequate for these signals, because it assumes that the complete analysis window is stationary. The Short-Time Fourier Transform (STFT) is a method to mitigate this problem by considering short segments of the signal at the time of computation instead of the whole signal at once. The STFT for a discrete speech signal \(x\[n\]\) may be formulated mathematically as follows:

$$X(m,k) = \Sigma[n=0 \text{ to } N-1]\, x[n+mH]w[n]e\text{^}(-j2\pi kn/N);$$

where the parameters define structural features of the analysis:

- $w[n]$ is the window function that tapers the segment of the signal
- H -the hop size between consecutive window placements
- m is the time index of the analysis frame
- N is the number of samples in the analysis window

The window function is an important factor in the quality of the time-frequency representation obtained. Popular window functions which may be used by speech analysis are:

Hamming window: $w[n] = 0.54 - 0.46\cos(2\pi n/N)$ Given by the equation I described first. Hanning window: $w[n] = 0.5(1 - \cos(2\pi n/N))$

It contains the Rectangular window $w[n] = 1$ as well, for all n.

Each window function has a different associated spectral behavior, i.e. there is a tradeoff between main side lobe width and side lobe attenuation. The side lobes of the Hamming window (-42 dB), which in the HammingвЂ™s window were already considered as good performance, particularly on speech processing applications, where the tradeoff between resolution and leakage is satisfactory.

The temporal resolution of the STFT is highly dependent on the duration of the window. Using longer windows (usually 20–30 ms for speech) allows more fine frequency resolution, which better discriminates close centre frequencies e.g., formants in vocalic sounds. On the other hand, shorter windows (5-10 ms) would have better time resolution, which is required for capturing fast acoustic changes in consonant manner productions such as plosive and affricate.

The hop size H is the stepping distance between analysis windows. Standard implementations employ 50-75% overlap for sufficient time resolution while remaining computationally feasible. This overlapping compensates for potential information loss at window borders and yields more continuous time evolution of spectral components. The time-frequency resolution trade-off follows from the Heisenberg-Gabor uncertainty principle, which establishes fundamental limits on simultaneous precision in time and frequency domains. In applications of speech analysis, this takes the form of:

- Wide windows → High frequency resolution and low time resolution
- Sharpe window → Good frequency resolution reducing (narrowing) in time.

Adaptive window duration method Adaptive window duration techniques are proposed for better voice analysis, in which window length depends on the signal properties. Spectral windows can be made longer to capture harmonic structure of the voiced segments and shorter to ensure the accurate temporal localization of unvoiced segments and transients.

Visualization The produced output of STFT is represented in the form of a time-frequency matrix and is often visualized using spectrograms, which are plot of time on the horizontal axis and frequency on the vertical axis and the color or grayscale intensity is proportional to the magnitude of the spectrum. This representation has been found to be useful in phonetic studies for visual identification of:

- Formant paths during veil production
- Explosive consonants: voice onset time
- On noise distribution in fricatives
- Temporal aspects of the suprasegmental structure

In contemporary STFT algorithms there are further enhancements added for speech analysis, such as:

- The window size is adjusted and based on pitch detection.
- Multi-resolution analysis using different window sizes
- High resolution spectral estimation via post-processing strategy
- Integration with perceptual frequency scales (Mel, Bark)

Spectrograms of English Speech Units Spectrogram Creation and Analysis

The spectrogram is an essential visualization tool in acoustic phonetics which summarizes speech signal over both time and frequency dimensions. This visual representation shows the magnitude squared of the Short-Time Fourier Transform ($|X(m,kmax)|^2$) as a function of the temporal progression on the horizontal axis and frequency components on the vertical axis. Gradient or gray-scale coloring indicates the strength of the spectral energy, where the hotter the color (or the darker the shade) the larger the energy accumulated.

Nowadays digital spectrograms are computed with combined settings for speech analysis. The frequency range typically extends from 0 to 8 kHz, which includes the most critical spectral information of human speech, with the ultrasonic part English 4 – – stok that cannot add much linguistic information being excluded. This distribution is an adequate approximation of:

- Changes in the fundamental frequency (50-300 Hz for most speakers)
- up to about 4 Kvz for vowel identification: Formant structures
- High frequency noise components essential for the discrimination of fricatives

An appropriate dynamic range of 30–50 dB was used such that stronger harmonic components and weaker noise components could be visualized simultaneously. What this span effectively shows:

- Well-defined, high prominence vowel formants and spectral peaks
- More subtle columnar auditory features such as aspiration noise
- Relative amplitude discrepancy among speech segments and simultaneously avoiding saturation effects that can mask interesting acoustic features.

Window duration is a key parameter in spectrogram computation. The 20-30 MS window convention represents an optimal compromise for English phoneme analysis by:

- Having enough frequency resolution to resolve formants (usually 300-400 Hz apart)
- Preserving approximately enough horizontal resolution to see the little lines I paint when I form consonants
- Capturing about 2-3 pitch cycles for normal voiced sounds

More advanced spectrogram engines employ other processing methods in order to facilitate phonetic interpretation. Pre-emphasis filters (e.g., 6 dB/octave) serve to equalize the natural spectral tilt in speech signal and to enhance its high-frequency components in order to obtain a clear fricative pattern. The logarithmic amplitude scale (in dB) has the advantage of showing weaker high frequency components better in relation to strong, low frequency energy. Perceptual Frequency Warping (Mel/bark scales) Some systems perform a warping of the frequency axis to better match the human auditory sensitivity.

Phoneme class must also be understood to interpret spectrographic presentations of various patterns. Vowels have relatively robust formant structures (with The Harmonic excitation), while the consonants have finer temporal-spectral organization. Transitional signals between the phonemes themselves are frequently important in phonemic perception, and nowhere is this better demonstrated than in stop-vowel contexts. The capacity to identify these complex patterns renders spectrograms indispensable in speech science research and clinical practice.

Modern DSP software can generate interactive spectrogram analysis with modifiable settings. This freedom allows to tune the representation for given desiderata, specifically:

- Analysis of pitch and harmonic structure (long windows) by narrowband revitalization
- Wideband analysis (short time windows) for consonant articulation research.
- Custom frequency range for specific use (e.g. child speech analysis)
- Time-scale and spectral representations are simultaneously derived for total signal evaluation

The quantifiable aspects of spectrographic analysis are conducive to exact measurements of acoustic-phonetic dimensions such as formant frequencies, bandwidth and trajectories. Index values obtained with such measurements make it possible to compare objectively between speakers, dialects, and languages, and serve as a material in numerous works within experimental phonetics and speech communication research.

Vowel Formants of English

Acoustic analysis of the produced vowels show differences in spectral characteristics, quantified as formant structures, corresponding to resonant frequencies of the vocal tract shape. These formants, listed in increasing order of frequency (F1, F2, and F3), are the main acoustic concomitants of the vowel quality and so represent measurable targets to categorize a vowel. Spectrographic analysis of English vowels: An acoustiophysiological correlation.

The English cardinal vowels have characteristic formant frequency patterns that are stabilized across normal individuals when spoken in the same context of other phonemes. F1 and F2 of the /i:/ that is the high front vowel as in “beet” have an obvious formant distribution with concentrated value of F1 near 300Hz and of F2 near 2300Hz. Such an acoustic pattern is due to a particular vocal tract constriction which requires tongue advancement and elevation, with lip spreading. The large separation of F1 and F2 (around 2000 Hz) is the basis of the bright quality of perception for this vowel.

By contrast, the low front vowel /æ/ (e.g., in "bat") has a strikingly different formant shape, with F1 raised to some 700 Hz, and F2 lowered to about 1800 Hz. This spectral characteristic corresponds to the open vocal tract shape and the front tongue position of this vowel category. The relatively high F1 is attributed to the large pharyngeal cavity volume and the F2 position indicates preserved front cavity resonance.

Back vowels exhibit inversion of formants, as in the high back vowel /u:/ (as in "boot"). This vowel has F1 around 350 Hz and F2 around 900 Hz, and the close formants are due to the combination of tongue backing with lip rounding, and the acoustic lengthening by the vocal tract. The small F2 value is particularly for closing back vowels distinctive in comparison to front vowels and adds to their perceived darkness.

Diphthongs exhibit more intricate spectrographic structures with dynamic formant transitions. For the diphthong /aɪ/ (as in “price”), the targeted formant movements are consistent with a situation in which F1 moves downward from an initial, fairly open position (≈700 Hz) to a more closed position (≈300 Hz) while F2 moves forward from a central position (≈1200 Hz) to a further fronted position (≈2000 Hz). Such formant paths indicate an articulatory transition from an open central to a close front vowel.

Acoustic information on the production of vowels can be found in the width of the formants. Those bandwidths typically the 50-150 Hz range for F1 and 60-200 Hz for F2 are indicative of the amount of damping of the vocal tract resonances. Findings regarding wide and narrow bandwidths, for example, have converged to indicate that wide bandwidths correspond to more centralized vowel productions or more breathy voice quality and narrow bandwidths with more peripherally produced vowel configurations.

The connections between formant frequencies and our perception of a vowel are based on well- known principles of psychoacoustics. In human listeners, F1 serves as the primary cue for vowel height (open vs. closed articulation) and F2 as the primary cue for vowel backness (front vs. back articulation). The less perceptually salient F3 frequencies are also used to distinguish between the rhotic and non-rhotic vowels of English.

Standardization procedures based on the formants of an individual's speech token are designed to control for individual differences in vocal-tract length whichaff ect absolute formant number. Sev- eral algorithms have been created taking raw formant frequencies and transforming them to correlate with normalized representation of phonemic system while account- ing for interspeaker variability. These tests are particularly important in vowel studies that involve different speakers and are often spoken by different age, gender, or ethnic groups.

Here are shown cross-dialectal contrasts in vowel formant structure. For example, the Northern Cities Vowel Shift shows measurable differences in the formant trajectory patterns of the vowels from those in Standard American English. These differences can be quantified for EGA by referring to ongoing sound changes and to regional dialect features; spectrography is the main documenting device.

Vowel formant analysis is used for a variety of clinical purposes, such as diagnosing and treating speech disorders. Atypical patterns of formants can signal articulatory imprecision, resonance impairment, or neuropathologies of speech motor control. In its clinical application, realtime formant displays are used as biofeedback to train vowel production to a speaker undergoing rehabilitation.

Acoustic Features of the English Consonants Plosive Consonants

The resulting acoustic realisation of the plosives (/p tm k bg/) has a three-part structure at the spectrographic level. The initial part is also a stop gap which is interpreted as a complete absence of acoustic energy during the articulatory closure. This period of silence lengthens or reduces with speaking rate and phonetic context, and its duration normally falls between 50–100 milliseconds. The following release burst emerges as a short (5-25 ms) wideband energy excitation over a wide frequency range, with the spectral properties depending on the place of articulation:

- Bilveolar stops (/p/, /b/) have broad, mid-frequency (500-1500 Hz) energy.

- Alveolar stops (/t/, /d/) exhibit more high-frequency concentration (1800–4000 Hz)
- Velar stops (/k/, /g/) show bimodal concentration in low (3000 Hz) frequency bands.

The principal acoustic cue which demarcates voiced and voiceless plosive pairs is voice onset time (VOT). Positive VOT values (10-30 ms for English voiced stops and 30-100 ms for voiceless stops) measure the degree to which the release of the burst precedes the vocal folds start vibrating. There is aperiodic energy in the spectrogram in the region of aspiration noise following release of the voiceless stop, which evolves into the regular form of the following vowels.

Fricative Consonants

Fricatives (/f/, /v/, /θ/, /ð/, /s/, /z/, /ʃ/, /ʒ/) generate continuous aperiodic noise by turbulent air flow across a point of constriction. Spectrographic Analysis The cyanide process can be distinguished as to several features:

Spectral Energy Distribution

- Labiodentals (/f/, /v/) exhibit a poor concentration of energy below 5000 Hz
- Dentals (/θ/, /ð/) exhibit a broad energy in the range from 3500 to 8000 Hz.
- The 4000 Hz threshold is more pronounced for the Alveolars (/s/, /z/)
- Palato-alveolars (/ʃ/, /ʒ/) have energy at 2000-4000 Hz.

Patterns of Duration and Intensity

- Unvoiced fricatives tend to be longer/louder than the voiced counterparts
- The differences in the spectral tilt reveal voice/voicelessness contrasts, with voiced fricatives having more low frequency energy.

Transitional Cues

- Formant transitions to surrounding vowels are valuable information on place of articulation.
- The F2 and F3 transitions especially seperate alveolar from palato-alveolar fricatives

Nasal Consonants

Nasal consonants (/m/, /n/, /ŋ/) show characteristic spectrographic features due to simultaneous oral closure and velopharyngeal opening:

1. **Formant Structure**

- Low-frequency formants (below 500 Hz) with bandwidths wider than vowels
- Characteristic nasal murmur showing stable formant patterns:

* /m/: F1 ≈ 250 Hz, F2 ≈ 1100 Hz, F3 ≈ 2400 Hz
* /n/: F1 ≈ 250 Hz, F2 ≈ 1700 Hz, F3 ≈ 2400 Hz
* /ŋ/: F1 ≈ 250 Hz, F2 ≈ 1300 Hz, F3 ≈ 2200 Hz

2. **Antiresonance Features**

- Nasal cavity-coupling induced spectral zeros
- Energy loss in laying out of selected frequency range (usually between 800 and 1500 Hz.

3. **Transitional Patterns**

- Fast transitions formant near the edge of the nasal-vowel segment
- Unique F2 and F3 trajectories for places of articulation

Affricates and Approximants

Affricates (/tʃ/, /dʒ/) contain plosive and fricative features in, for example:

- Initial stop gap and burst
- Long frication noise (30-80 ms) with palato-alveolar spectral character
- Clean formant transitions to the subsequent vowels Approximants (/l/, /r/, /w/, /j/) are:
- Vowel-like formant structures with more constriction
- Quick formant change which reflects articulatory gesture.
- F3 patterns that discriminate rhotics (/r/) from other liquids

Acoustic analysis of English consonants shows that clear relations exist between articulatory states and the spectrum/time properties of corresponding sounds. These patterns form the basis for theoretical phonetic studies and for practical applications in speech technology and clinical phonetics.

Applications and Implications Applications of Linguistic Research

The developments in experimental phonetics due to spectrographic analysis, and quantitative research on speech production. Through systematic formant distribution analysis this method has become invaluable in studies of dialect variation. Through comparison of the vowel spaces of regional accents of English, different levels of acoustic and articulatory detail can be recorded by researchers, which may not be detectable to non-specialist listeners. For example, spectral evidence has demonstrated the Northern Cities Vowel Shift in terms of directly observable formant dynamics across time.

The phonologization of historical sound changes is a domain where spectrographic analysis has made considerable advances. By studying archival recordings of speakers across generations, linguists can trace subtle changes in patterns of articulation that reflect the continuing evolution of language. Geyer's methods have been used to establish phonological differences, such as the neologism slut as a variant of the Standard English slot, either in recent result or in process of evolution (for Geyer, both), between mid-20th century British and American English, i.e. showing that "the low back vowel merger is taking place in one dialect but not in the other". Spectrograms are the graphic representation of a spectrogram, and work as an encoding of speech to visual representation complementary, to some extent, to traditional phonetic transcription.

In the area of child language acquisition, spectrographic analysis is used to study how speech production skills develop. Longitudinal studies of formant frequency in children\u2019s vowel production show these to be acquired gradually as articulatory control is refined with phonological development. The method captures exactly how children’s vowel spaces grow and settle, and reveals the nature of the process by which their realization of phonemic contrasts is acquired.

Furthermore, since spectrograms can reveal abnormal developmental patterns which may indicate or lead to speech or language pathology.

Phonetic studies of non-native speech learning use spectrogram analysis in the attempt to quantify the degree to which learners of a second language succeed in approximating the sounds of the target language. When measured against formant patterns of native speakers, particular articulatory obstacles may emerge that can help to track the phonetic learning process. This will have implications in second language teaching approaches as well as accent modification treatments.

Sociolinguistic projects also employ spectrographic measurements to analyze the relationship between social characteristics and speech modes. Acoustic analysis of vowel formants has identified consistent differences across age, gender, socioeconomic group, and ethnic identification. These results add insight about how linguistic variation serves as an index of social identity in speech communities.

Spectrographic analysis of fundamental frequencies is useful in prosodic studies, as well as timing. This visualization of pitch contours and temporal features allows us to investigate properties of intonational phonology, stress and rhythm in varieties of English. It has brought us closer to understanding how prosodic factors shape meaning in speech communication.

Instrumental phonetics has been further developed by the combination of spectrographic analysis with electropalatography and ultrasound tongue imaging. These multimodal methods afford rich information on the mutual mapping between the articulatory gestures and their corresponding acoustic consequences and thus contribute to the further development of theories of speech production. This combined methodology has been especially useful for observing coarticulation effects and connected speech phenomena.

Implementing Speech Technology

Audio-based signal processing via Fourier transforms is now a fundamental building block of most modern speech technology systems, with remarkable improvements made in multiple application areas. The computation of speech signal into frequency-domain representations and devices that respond to these representations have become an integral part of machine processing and analysis and has led to the development of a whole spectrum of technology that has had a significant impact in the area of human-computer interaction and assistive communication devices.

Mel-Frequency Cepstral Coefficients (MFCCs) are one of the most popular Fourier derived features used for automatic speech recognition systems. This method uses a series of processing stages, starting from Fourier analysis of windowed speech segments, then mapping of the frequency to the perceptually important Mel frequency scale. The rerulting cepstral coefficients are effective in capturing the spectral envelope structure in a pitch-invariant manner, and thus prove very effective in speaker-independent tasks. Current implementations commonly use 12-20 MFCC features with their delta and delta-delta features to model static and dynamic spectral features.

Formant tracking algorithms are of prime importance in text to speech synthesis systems to estimate the resonances of the vocal tract to produce natural sounding artificial speech. The extraction of formant can be considered as the application of Fourier-based spectral analysis, while the more recent schemes might be inspired from Fourier-like principles (like in the Linear predictive coding (LPC)). They use formants cues (frequencies and bandwidths) to the control the parameters of synthetic speech, which facilitates an arbitrary movement of vowel quality and a smooth transition of formant frequencies across phoneme boundaries. The combination of Fourier-based formant synthesis with concatenative/pseudorandom techniques has led to much improved naturalness and intelligibility of synthetic speech output.

Indeed, in hearing aids, the algorithms of noise reduction are essentially based on Fourier analysis to discriminate between the speech and the ambiant noise. Such systems usually use short-time Fourier transform for the decomposition of the acoustic signal into time-frequency cells, and then statistically estimate the noise and speech components. Modern digital hearing aids apply spectral subtraction techniques, which estimate and suppress noise-dominant frequency regions, in order to preserve speech-dominant parts. Recent advances combine machine learning approaches to Fourier-based features to improve listening conditions for speech in difficult environments.

Fourier based features are widely used by speaker recognition/verification system applied to voice biometrics. Speaker-dependent patterns extracted by Fourier analysis in the spectral domain are relatively invariant across utterances. One popular class of those systems integrates traditional Fourier-based features with more recent neural network approaches, keeping the base frequency- domain analysis but enhancing discrimination interpretation through deep-learning architectures.

Fourier and Fourier-related transforms are widely used in speech coding and compression algorithms for efficient representation of signals for transmission and storage purposes. The CELP coders (code-excited linear prediction coders) which are used in many voice communication standards also include a spectral analysis step which is based on the Fourier transformation. These methods achieve large data reductions while keeping to within acceptable limits the quality of speech, by preserving important perceptual spectral features determined through the use of a Fourier analysis.

Speech-based emotion recognition has become a significant application domain for Fourier features. The spectral properties of speech, in particular in higher frequency bands, have been shown to be associated with different emotional states. Emotion recognition systems extract Fourier-based spectral features in conjunction with prosodic information, which makes systems applicable to various domains including customer service analytics, mental health monitoring and human – computer interaction systems.

Real-time spoken language processing systems used in language learning utilize Fourier analysis to give instantaneous feedback on the accuracy of pronunciation. By comparing their vowel formant patterns and consonant spectral characteristics with those of native speakers, such systems can determine which articulation errors are being made and how correction is to be achieved. The graphical expression of the Fourier-based spectral analysis has been found to be a highly effective tool when incorporated into interactive learning tools.

The further evolution of such Fourier-based approaches in speech technology is indicative of their sustained significance, despite the age of such methods relative to many other approaches. The work of Fourier provided the mathematical underpinning for much of the subsequent development in speech processing and have been used up until now as key time-frequency representations (TFRs) in many hybrid systems that span classical signal processing to state-of-the-art machine learning.

Resources: Clinical Applications in Speech-Language Pathology

Assessment via spectrographic analysis has become an integral part of modern speech-language pathology. The visualization of speech acoustics offers an objective data pattern, which can augment the conventional treatment though perceptual methods, and can also help provide measurable proof of the effects of various communication disorders. Clinical examples include assessment, intervention, and progress monitoring for a variety of speech and voice disorders.

Spectrographic analysis contributes to the diagnosis of articulation disorder when speech sound errors are more easily detected and described. The method provides differentiation between phonological process and motor speech disorder by eliciting characteristic acoustic profiles for each. Spectrograms also allow for visualization of small articulation differences, which may not be easy to discriminate by aural examination alone, such as minor modifications to place or manner of articulation. The value of examining voice onset time in plosive production is especially evident when assessing speechmotor control difficulties that are related to timing of movements.

Visual feedback therapy uses spectrographic representations as biofeedback for changing the speech sound. Real-time spectrograms offer the client direct visual feedback on their speech production, enabling them to monitor themselves in real time with respect to the target sounds. This method has been reported effective for residual articulation errors, especially those consonants with difficult-to-hear frequency features as /r/ and /s/. The visual feedback-based paradigm complements conventional auditory-driven therapy by engaging another sensory modality for learning and self-monitoring.

Spectrographic analyses are used to assess a number of acoustic parameters of vocal production. Abnormal voice qualities such as hoarseness, breathiness, and strain exhibit unique patterns on spectrograms. Jitter and shimmer extracted from spectrographic analysis measure the irregularity of vocal fold vibration related to laryngeal pathologies. The procedure offers an objective record of vocal parameters pre- and postmedical or behavioral intervention and acts as a baseline measure for treatment outcome.

Spectrographic analyses aid in the diagnosis and treatment of childhood apraxia of speech (CAS) by identifying inconsistent vowels errors and anomalous prosodic patterns. The comprehensive acoustic analysis contributes to the distinction between CAS and other speech sound disorders and to the development of specific clinical management protocols. Formant transitions analysis is especially helpful to assess sequencing problems in motor speech programming.

Treatment of a disorder of fluency includes the study of the acoustic correlates of moments of stuttering using spectrography. The method enables detection of subtle prepatory movements and concomitant vocal tract readjusting which often being overt dysflueces. Spectrographic feedback is a promising addition to comprehensive therapy for stuttering, especially as it facilitates controlled phonation and speech initiation.

Spectrographic analysis, which identifies abnormal formant patterning related to VPD, aids in the assessment of resonance disorder. The method aids in distinguishing hypernasality, hyponasality, and cul-de-sac resonance according to their characteristic spectral features. In addition, nasal consonant spectrograms are particularly helpful in evaluating the adequacy of velopharyngeal closure.

Rehabilitation of hearing impairment uses spectrographic images to assist in speech sound development of children with cochlear implant or hearing aids. The visual display of acoustic targets offsets the individuals' impaired discriminative abilities, and as a result the correct patterns of articulation are acquired. Spectrograms are useful for the instruction of speech sound contrasts that are difficult to hear in a student's worn hearing losses alone.

Management of neurogenic speech disorders includes the use of spectrographic analysis to describe the acoustic effects resulting from central nervous system diseases. Dysarthric subtypes exhibit unique spectrographic signatures and can be used to guide differential diagnosis and treatment. It also offers a quantitative way to measure improvement or decline of speech in a progressive neuromuscular disease and could be a sensitive marker of disease progression or response to treatment.

Spectrographic analysis has increased the evidence for speech-language pathology practice through quantifiable acoustic measurements, as opposed to purely perceptual judgments. The inclusion of this technology into clinical practice is still developing, and current research is continuing to examine new roles in assessment protocols and intervention techniques. Spectrographic displays provide an excellent visual resource which increase the client's comprehension of speech production targets and improve the accuracy of self-monitoring of production during therapy.

Conclusion

The introduction of Fourier analysis and spectrographic techniques has revolutionized the scientific study of English speech sounds and made both methods standard tools for present-day phonetic research. The mathematically rigorous model of Fourier transforms allows for quanti ed examination of the sound spectrogram, to reveal characteristics of speech sounds that might otherwise remain out of reach of pure auditory analysis. This method of analysis has allowed the most thorough examination yet of spectral and temporal features that might differentiate the English phonemes.

The visualization of spectrographic representations is particularly useful for the study of phonetics as detailed scrutiny of both segmental and suprasegmental properties is made possible. The time- frequency representations created by well controlled use of digital signal processing methods faithfully reflect much of the dynamism of the speech production process. These graphical displays have in turn become routine instruments for the analysis of formant structure, voice source information, and the timing of articulatory gestures in various speech sounds.

That Fourier analysis, with its properties of orthogonality and projection, continues to play an important role in speech science is evident from its ready acceptance in diverse domains of investigation. Beginning at the interface of theory and practice these approaches offer a shared reconceptualisation that spans the continuum of the theoretical and the applied. This calculation offers the higher levels of reliability and validity in response to experimental verification in scientific research, due to the mathematical rigor derived from Fourier analysis.

Current state of the art in digital signal processing has extended the use of spectrography by keeping Klein's even if principle based on Fourier transforms. Contemporary implementations use traditional Fourier analysis in conjunction with advanced computational methods, which has increased the scope of phonetic phenomena that can be studied in a systematic way. Such advances have allowed more accurate acoustic measurement and clear representation of subtle characteristics of speech that contribute to phonemic contrasts.

Peace to the memory of Pierre Simon Laplace The synthesis of Fourier analysis with the new technology of Speech Science offers exciting possibilities for gaining new knowledge about the human production and perception of speech. Although there are emerging alternative analytical methods the fundamental importance of

Fourier-based spectrography is not jeopardized by the new methods since the theoretical foundation is sound and the method has shown practical value. The fact that the method is able to span the divide between acoustic signal and linguistic description is one reason why it has retained its significance with regard to phonetic research.

It is expected the future of speech science will extend rather than obviate the seminal input of Fourier analysis. The adaptability of the method to new research questions and to technological applications justifies its continued relevance in the ever changing scientific context. As the discipline continues to move toward more advanced models of speech production and perception, the information provided by spectrograms will continue to be important in testing theoretical constructs and in suggesting that certain physical units are more potentially motivated in the language than others.

One should not undervalue the pedagogical aspect of spectrographic analysis: it offers a way not only to make theory graphic and visible, but also to make visible an abstract phonic reality. That has been particularly important in training a generation of new phoneticians and speech scientists and in promoting a better understanding of the link between articulatory processes and the acoustic results of their production. The pedagogical implications serve to emphasize, additionally, the significance of the technique in the discipline.

In conclusion, Fourier analysis and spectrographic methodology have become valuable tools in the study of the speech sciences and have provided the base for a variety of basic and applied advances. Their further development and use will undoubtedly be beneficial in the future to advance our knowledge of English phonetics and the field of speech science more generally. The flexibility and robustness of these approaches lead us to expect that they will continue to guide efforts in the field as new fronts in research and technology open.

References

- Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* [Computer software]. http://www.praat.org/
- Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech* (2nd ed.). Singular.
- Ladefoged, P., & Johnson, K. (2015). *A course in phonetics* (7th ed.). Cengage Learning.
- Oppenheim, A. V., & Schafer, R. W. (2010). *Discrete-time signal processing* (3rd ed.). Prentice Hall.
- Rabiner, L. R., & Schafer, R. W. (2007). *Theory and applications of digital speech processing*. Pearson.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928. https://doi.org/10.1126/science.274.5294.1926