

The Importance of Clarity in Speech and Its Acoustic Properties

أهمية الوضوح في الكلام وخصائصه الصوتية

DAOUD Messaoud^{1*},

¹Full affiliation University of Algiers 2 – Abou El Kacem Saadallah (Algeria),

messaoud.daoud@univ-alger2.dz

<https://orcid.org/0009-0000-6157-2908>

Received: 12-07-2025 Accepted: 17-04-2026 Published: 2026-05-19

Abstract :

This research aims to highlight the importance of clarity in speech by investigating the acoustic properties that underpin intelligible communication. The study analyzes these properties—such as fundamental frequency, formants, and amplitude—to assess how they contribute to perceptual clarity, focusing on the central research question: What are the key acoustic characteristics that determine speech clarity, and how do they interact to ensure effective transmission of linguistic content?

The study yielded several significant findings, most notably that speech clarity is directly dependent on the precise configuration of formant structures and the stability of fundamental frequency. Moreover, temporal features such as transition duration and voice onset time play a critical role in distinguishing phonemes, thereby affecting overall intelligibility.

Keywords: Speech clarity, acoustic properties, fundamental frequency, formants, intelligibility.

ملخص:

يهدف هذا البحث إلى إبراز أهمية الوضوح في الكلام من خلال دراسة الخصائص الصوتية التي تقوم عليها عملية التواصل المفهوم. وتحلل الدراسة هذه الخصائص — مثل التردد الأساسي، والصيغ (الفورمانتات)، والشدة الصوتية — من أجل تقييم مدى إسهامها في تحقيق الوضوح الإدراكي، مع التركيز على السؤال البحثي المركزي: ما الخصائص الصوتية الأساسية التي تحدد وضوح الكلام، وكيف تتفاعل هذه الخصائص لضمان نقل فعال للمحتوى اللغوي؟

وقد أسفرت الدراسة عن عدة نتائج مهمة، أبرزها أن وضوح الكلام يعتمد بشكل مباشر على الضبط الدقيق لبنية الصيغ الصوتية واستقرار التردد الأساسي. إضافة إلى ذلك، تلعب الخصائص الزمنية مثل مدة الانتقال وزمن بدء التصويت دورًا حاسمًا في تمييز الفونيمات، مما يؤثر في مستوى الفهم العام للكلام.

كلمات مفتاحية: وضوح الكلام، الخصائص الصوتية، التردد الأساسي، البواني، القابلية للفهم.

1. Introduction:

Language is one of the most fundamental means of social interaction, given its vital role in self-expression and understanding others. It is directly linked to emotional, cognitive, and intellectual development, as it enables connection between individuals and groups. A speaker's ability to convey meaning through language relies on the clarity of the speech signal, a process that requires specific acoustic properties to ensure effective transmission.

Speech (voice) is considered a core component of spoken language and plays a crucial role in how meaning is perceived by the listener, who processes language primarily through sound. However, several factors may compromise

* Corresponding Author.

speech clarity, including environmental interference or technical distortions in transmission channels, leading to reduced intelligibility. This raises the need to investigate the acoustic properties of linguistic sound and the extent to which they contribute to clarity. Accordingly, this study seeks to address the following research question: What are the key acoustic characteristics that determine speech clarity, and how do various interferences affect them?

The primary objectives of this study can be summarized as follows:

- To highlight the importance of speech clarity in effective listener perception.
- To underscore the significance of employing acoustic phonetics in the analysis and processing of speech sounds for clarity enhancement.
- To utilize modern software tools in the analysis and assessment of speech signals.

The significance of this study lies in the integration of modern technological tools—essential instruments of the contemporary era—in the processing of speech, which represents a key component of spoken language. By conducting an acoustic analysis of speech, the study aims to extract its defining features, which in turn provide valuable insights into its clarity. Among the primary parameters used to assess speech clarity are fundamental frequency, formants, and intensity.

This study adopts a descriptive-analytical approach aimed at presenting the fundamental aspects of the acoustic properties of speech sounds. It involves the collection of various data—specifically acoustic features—and their subsequent analysis to extract key observations and findings. The analytical component focuses on examining and interpreting speech signals using specialized instruments and tools designed for acoustic analysis.

2. Linguistic Sound and Its Acoustic Properties

2.1 Definition of Sound:

Sound is fundamentally defined as an auditory sensation that arises from the propagation of mechanical waves originating from the vibration of a physical object or source. When an object vibrates, it imparts kinetic energy to the surrounding particles, creating alternating zones of compression and rarefaction that travel outward through an elastic medium. As such, sound constitutes an external physical disturbance—a form of mechanical energy transfer—that exerts influence on all matter encountered along its path. This influence is most evident in gases such as air, where the particles are displaced from their equilibrium positions and set into oscillatory motion as a direct consequence of the vibration emitted by the sound source. The resulting chain of particle interactions enables the wave to propagate, carrying acoustic energy from the source to a receiver, where it is ultimately interpreted by the auditory system as sound. (Khaoula, 2006, p. 45). As sound propagates through a medium, each

individual air particle—having been disturbed from its resting state—transfers its kinetic energy and oscillatory motion to the adjacent particle through a series of elastic interactions. This process of sequential energy transfer repeats continuously along the propagation path, giving rise to a chain of successive sound oscillations in which each oscillatory cycle is directly derived from and dependent upon the one immediately preceding it. The oscillations do not travel as discrete particles moving from source to receiver; rather, it is the pattern of disturbance that moves forward, passing from one particle to the next in a coordinated fashion. This orderly progression of mechanical disturbances—characterized by alternating compressions and rarefactions traveling through the transmission medium—is formally defined as a sound wave. The sound wave thus represents the collective behavior of countless particles oscillating in phase, forming a coherent structure that carries acoustic energy across distances. (Omar, 1997, p. 27).

2.2 Characteristics of Sound:

a) Displacement:

Displacement, in the context of sound wave propagation, describes the measurable distance—expressed in units of length—that a particle moves away from its original resting position, also known as its equilibrium point, as a result of the mechanical vibrations generated by a sound source. This movement is not a permanent relocation but rather a temporary, oscillatory shift that varies continuously over time as the wave passes through the medium. The magnitude of displacement at any given moment reflects the instantaneous position of the particle relative to its point of rest, and it serves as a fundamental parameter for characterizing the amplitude and energy of the acoustic wave. As the vibration intensifies, the displacement range expands, while weaker vibrations produce more limited deviations from equilibrium. (Saad, 2005, p. 29).

Greater particle displacement produces stronger acoustic pressure, enhancing the audibility of subtle articulatory details essential for distinguishing phonemes and ensuring speech clarity.

b) Vibration:

Vibration is the repetitive, oscillatory motion that a physical body performs when it is disturbed from a state of equilibrium in order to generate sound. The process unfolds in a precise sequence: the object departs from its resting position—the point of equilibrium—and travels in one direction until it reaches a first extreme point, where it momentarily comes to a temporary stop. It then reverses direction, moving back through the equilibrium position and continuing onward to a second extreme point on the opposite side, where it again halts briefly. After this second pause, it reverses once more and returns to the original starting position. This entire back-and-forth trajectory constitutes one complete cycle of movement, which is technically termed a vibration or an oscillation. When such cycles are repeated in a regular, periodic fashion—each cycle having

the same duration and pattern under stable conditions—the resulting phenomenon is referred to as acoustic vibration or sound oscillation. This periodic motion is fundamental to the generation of sound waves, as it creates the alternating compressions and rarefactions in the surrounding medium that propagate as audible sound. (Mahmoud & Abdullah, 2009, p. 134).

Regular, stable vibration of the vocal folds produces a consistent fundamental frequency, providing the acoustic foundation upon which clear pitch contours and phonemic distinctions are reliably perceived.

c) Amplitude:

Amplitude is defined as the maximum extent of displacement that a particle achieves from its equilibrium or resting position during one complete cycle of oscillation. It quantifies the distance between the point of rest and the farthest point reached by the moving particle, thereby representing the peak magnitude of vibratory motion. This distance is not constant; rather, it varies in direct proportion to the range of motion exhibited by the vibrating body itself, which is determined by how far the source moves from its equilibrium to its extreme endpoints. Consequently, amplitude increases or decreases in accordance with the intensity of the force that initiates and sustains the vibration of the sound source: a more forceful input imparts greater energy, resulting in larger amplitude, while a weaker force yields smaller amplitude. These variations occur within the transmission medium, where the sound wave correspondingly expands or contracts in scale, reflecting the strength of the underlying vibration.

Amplitude has a direct and pronounced effect on the human auditory system. When sound waves reach the ear, the amplitude of the oscillations determines the perceived loudness. Specifically, the greater the amplitude, the stronger and more intense the sound is perceived by the ear, as higher-amplitude waves displace greater volumes of air and exert more pressure on the tympanic membrane. Conversely, lower amplitudes result in weaker auditory sensations.

The Relationship Between Sound Intensity and Amplitude:

a. A loud sound is produced when the vibratory motion exhibits a large amplitude, meaning the source oscillates over a wide range, imparting substantial energy to the medium and generating a high-intensity acoustic wave that the auditory system interprets as powerful and resonant.

b. A soft sound, in contrast, results from a small amplitude of vibratory motion, where the source moves within a narrow range, transferring minimal energy to the medium and producing a low-intensity wave that the ear perceives as faint, quiet, or subdued. (NourAl-Din, 1992, p. 105)

Adequate amplitude ensures that subtle acoustic contrasts, such as voicing and formant transitions, remain perceptible, directly supporting the intelligibility and clarity of spoken language.

d) Frequency:

Frequency is defined as the number of complete oscillations or cycles that a particle executes within a time span of one second as sound propagates through a medium. It serves as a fundamental measure of the rate at which the repetitive vibratory motion occurs, distinguishing one sound from another in terms of pitch. Different sounds exhibit varying frequencies because each vibrating object possesses its own characteristic natural frequency—the frequency at which it tends to oscillate when set into motion. This natural frequency is not fixed universally but rather changes according to specific physical factors inherent to the vibrating body, among which mass and length are particularly significant.

The relationship between frequency and physical properties depends on the type of vibratory system under consideration. In the case of strings—such as those found in stringed musical instruments—frequency is primarily influenced by the degree of tension applied to the string; increased tension raises the frequency, while decreased tension lowers it. Additionally, factors such as string length and linear density also play determining roles. In the case of resonating cavities—including air columns in wind instruments or the vocal tract in speech production—frequency is affected by mass, shape, and volume of the cavity, as these parameters define the resonant modes and the wavelengths that can be sustained. Frequency is measured in hertz (Hz), a unit that signifies one cycle per second, named after Heinrich Hertz, the German physicist who made pioneering contributions to the study of wave phenomena. (Mansour, 2001, p. 105).

The frequency range of linguistic sound waves generally extends from a lower limit of around 50 hertz to an upper limit of approximately 10,000 hertz (Mansour, 2001, p. 110).

Sound is intrinsically linked to frequency in the following way: vibrations occurring at low frequencies generate a deep, low-pitched auditory perception, whereas vibrations occurring at high frequencies produce a sharp, high-pitched auditory sensation. (NourAl-Din, 1992, p. 105).

Precise frequency control ensures accurate pitch perception and distinguishes phonemes, particularly vowels and consonants, thereby playing a fundamental role in maintaining overall speech clarity.

e) **Time:**

In the analysis of acoustic signals, the temporal measurement of sound waves is commonly expressed in milliseconds, a unit of time equivalent to one thousandth of a second, allowing precise quantification of duration, period, and timing events. (Mansour, 2001, p. 105) Within the framework of spectrographic analysis conducted using a sonograph, the dimension of time is conventionally represented along the horizontal axis of the resulting visual display.

Time constitutes a fundamental dimension in the identification and description of speech sounds across various phonetic contexts, playing an indispensable role in capturing the dynamic nature of articulation. Its significance becomes especially pronounced when analyzing sounds that are characterized by

a complete and transient closure of the vocal tract, which is subsequently followed by a sudden and abrupt release of built-up air pressure. This articulatory sequence yields what is acoustically referred to as a plosive burst, a distinctive transient event that marks the release phase of stop consonants. Such temporal features are clearly observable in the articulation of the consonant /d/, where the precise duration of closure, the timing of the burst, and the voice onset time collectively contribute to its perceptual identity.

Furthermore, time plays an indispensable role in measuring the transition duration that occurs when moving from one speech sound to another, thereby offering valuable insight into the phenomenon of coarticulation as well as the broader temporal dynamics that govern the production of connected speech. (Mahmoud & Abdullah, 2009, p. 167, 168).

Precise temporal control of closure duration, voice onset time, and transition intervals ensures accurate phoneme distinction, directly contributing to the intelligibility and clarity of spoken language.

f) Pitch (Fundamental Frequency):

Commonly known as pitch or, in more technical terms, sonic sharpness, this particular property is described by physicists as the perceptual auditory characteristic that allows the human ear to differentiate between sounds or tones, categorizing them as either high in pitch—often described as sharp—or low in pitch—frequently characterized as deep. (NourAl-Din, 1992, p. 154).

From a physical standpoint, the perceptual difference between sounds characterized as high-pitched and those described as low-pitched is directly attributable to the frequency of the sound wave involved. This frequency, in the context of complex acoustic waves, corresponds to what is known as the fundamental frequency—the lowest and most dominant frequency component that determines the overall pitch sensation. When the frequency is higher, the listener perceives a correspondingly higher pitch; when the frequency is lower, the perceived pitch becomes deeper, more guttural, or more resonant in quality. (Omar, 1997, p. 30).

In the context of human voice production, the frequency of the emitted sound is determined by the rate at which the vocal folds vibrate. When the vocal folds oscillate more rapidly, the resulting pitch is perceived as higher; when they vibrate more slowly, the pitch is lower. The speed of vocal fold vibration is not uniform across individuals but varies according to factors such as the size and tension of the folds. Generally, longer and thicker folds tend to vibrate more slowly, producing a lower pitch, while tighter folds vibrate faster, yielding a higher pitch. (Mahmoud & Abdullah, 2009, p. 154).

Consistent pitch variation conveys intonation and emotional nuance, enabling listeners to parse meaning accurately and enhancing the overall clarity of spoken messages.

g) Spectral Representation of Sounds:

In acoustic spectrography, the spectral bands that manifest as horizontal dark striations—specifically designated as the first formant (F1), second formant (F2), and third formant (F3)—play an essential role in the identification of voiced sounds. These are sounds produced with periodic vibration of the vocal folds, and the formant frequencies reflect the resonant characteristics of the vocal tract. Among the voiced sounds that rely on formant structure for their perceptual identity are vowel sounds, nasal consonants such as /n/ and /m/, lateral consonants like /l/, and non-turbulent continuants (including approximants and other sonorants). In addition to its utility for voiced segments, spectrographic analysis also contributes to the examination and detailed description of certain fricative sounds, for instance /h/ and /x/ (the latter representing a voiceless uvular fricative), whose spectral properties differ markedly from those of voiced sounds.

The accompanying figure presents a spectrogram of the vowels /i/ (kasra) and /u/ (damma), clearly illustrating the formation of formant bands. The key spectral features observable in such a display include the steady-state regions where formants remain stable, the transitions that signal coarticulatory effects, and the relative spacing between formants, which distinguishes one vowel quality from another.

Beyond formant patterns, some linguistic sounds appear in spectrograms as irregular distributions of energy—manifesting as scattered dots or fragmented vertical striations. This visual representation proves especially valuable for analyzing fricative sounds, where aperiodic noise occupies specific frequency ranges, as well as for examining the transient bursts associated with plosive consonants and the rapid, multiple closures characteristic of trilled consonants such as /r/. By integrating these diverse acoustic cues, spectrographic analysis provides a comprehensive framework for investigating the phonetic properties of speech. (Mahmoud & Abdullah, 2009, p. 172).

Distinct formant patterns enable accurate vowel and consonant identification, while spectrographic visualization reveals critical acoustic cues essential for achieving and assessing speech clarity.

3. Conclusion:

This study set out to investigate the relationship between speech clarity and the acoustic properties of linguistic sounds, focusing on how clarity—rooted in acoustic parameters such as fundamental frequency, formants, and amplitude—serves as a cornerstone for effective spoken communication. Through a descriptive-analytical approach and the use of spectrographic analysis, the research examined the acoustic features of speech and their role in determining intelligibility.

The main findings of the study can be summarized as follows:

- Clear speech relies on the precise articulation of formant structures (F1, F2, F3), stable fundamental frequency, and appropriate amplitude modulation. Any

disruption in these parameters, whether due to physiological factors or environmental interference, reduces intelligibility.

- External noise—both environmental and channel-related—distorts the spectral and temporal characteristics of speech. It obscures formant transitions, masks high-frequency fricative energy, and introduces irregularity in pitch contour, leading to perceptual ambiguity.
- Temporal aspects such as closure duration, voice onset time, and transition intervals between phonemes significantly impact the accurate identification of consonants and vowels. Spectrographic analysis reveals that even minimal timing deviations can alter the perceived identity of a sound.
- Modern acoustic analysis tools provide detailed visual representations of both spectral (formants, noise bands) and temporal (bursts, transitions) features, enabling precise evaluation of speech clarity and offering a reliable method for diagnosing and remediating clarity issues.

Based on these findings, the following recommendations are proposed:

- Clinicians, language educators, and communication specialists should adopt spectrographic feedback to help speakers identify and correct clarity-impairing patterns, such as insufficient formant differentiation or irregular pitch modulation.
- In technological applications (e.g., telecommunication, public address systems), emphasis should be placed on preserving the spectral and temporal integrity of speech signals through noise-reduction algorithms and bandwidth optimization.
- Because speech clarity strategies may differ across languages, future studies should extend acoustic analysis to a wider range of phonetic inventories and prosodic structures to establish universal and language-specific benchmarks.
- Raising awareness among speech-language pathologists, educators, and learners about the acoustic underpinnings of clarity will foster more effective teaching and remediation strategies.

In conclusion, clarity in speech is not merely a matter of articulation but is fundamentally grounded in the acoustic properties of the signal. A detailed understanding of these properties, combined with systematic spectrographic analysis, provides a robust framework for improving both natural and technologically mediated communication.

4. Bibliographic /references and sources:

- Al-Ghamdi, M. (2001). Arabic Phonetics. Maktabat al-Tawbah, Saudi Arabia.
Allam, A. & Mahmud, A. (2009). Phonetics. Maktabat al-Rushd, Saudi Arabia.
Maslough, S. (2005). Study of Hearing and Speech. Alam al-Kutub, Egypt.
Nour al-Din, I. (1992). Linguistic Phonetics. Dar al-Fikr al-Lubnani, Lebanon.
Omar, A. M. (1997). Linguistic Sound. Alam al-Kutub, Egypt.
Talib al-Ibrahimi, K. (2006). Principles of Linguistics. Dar al-Qasaba li-l-Nashr, Algeria.