Articulation Index Time Line

By
Christine Rankovic, PhD
Founder/Scientist
Articulation Incorporated

The first significant effort to understand the impact of various distortions on speech intelligibility was made by AT&T's Western Electric Research which was renamed Bell Telephone Laboratories (BTL or Bell Labs) in 1925. The telephone company supported a comprehensive internal research program during the most of the twentieth century whose original goal was to improve the clarity of telephone speech. AT&T's strong commitment to research may have been motivated by their desire to maintain an agreement with the U.S. government which granted them what was essentially a monopoly on the nation's telephone system from 1913 until 1985. The Bell Labs research program included an acoustical research group staffed by top scientists who designed original equipment and methods for sound measurement, generation and control; developed procedures for conducting laboratory experiments to determine the fundamental constants associated with speech and hearing; carried out speech recognition tests to evaluate telephone components; and conducted field surveys of real-life listening conditions and communication behavior. Their work was vital to creating the best telephone system in the world. They used their knowledge to invent the condenser microphone, motion-picture sound, stereo recording/reproduction, the audiometer, the first vacuum tube hearing aid, and later, the transistor-based hearing aid (the first application of the transistor, a Bell Labs invention), and more. Their empirical findings and theoretical conclusions—most of which were published in prestigious scientific journals and in books—are foundational and persist as definitive works in the speech and hearing sciences, although the Bell Labs contributions are seldom given their due credit.

The telephone company wanted high quality speech transmission at a reasonable cost. They realized that knowing the quality of a communication system in advance of implementation would prevent costly errors. *Harvey Fletcher* was among the scientists working on this and his solution was the *articulation index* (*AI*). The *AI* is a metric that predicts the capacity of a communication system to transmit intelligible speech. The *AI* is calculated using acoustical and/or electroacoustical descriptors of the communication system, namely, the frequency response, gain, speech intensity level, and the spectrum of any interfering noise. The *AI* calculation was intended to replace listener tests because they require inordinate effort to collect reliable and valid scores.

A physicist by training, Fletcher was employed by the telephone company from 1916 to 1948, and served as the director of acoustical research from 1928 to 1948. His *AI* work first appeared in a 1921 internal report. Fletcher oversaw years of evaluations and revisions as his group's knowledge and database grew, and this new-found information was incorporated into the *AI* calculation. The final finely-tuned version of the *AI*—first published in 1950 (with Galt)—is the basis for **Articulation Inc**'s Interactive *AI*™.

CRANDALL (1916). Fletcher's *AI* work was inspired by his mentor at the telephone company, I.B. Crandall. In a 1916 AT&T internal report, Crandall described a calculation for predicting *articulation scores* using frequency response measurements. An *articulation score* is a particular type of intelligibility test percent-correct score in which a list of nonsense syllables is read to listeners for identification, listeners record what they hear, and then each speech sound (or each syllable) is scored as correct or incorrect. Earlier investigations found that scores collected using lists of meaningful words or sentences were unreliable due to psychological factors such as practice, memory, familiarity, and linguistic context effects. The superiority of nonsense syllables was demonstrated in experiments in which syllables were spoken by trained talkers to trained listeners, and the group average percent-correct was taken as the articulation score.

Crandall's was the first attempt to calculate articulation scores. He examined articulation scores from experiments in which cutoff frequencies of high- and low-pass filters were systematically varied, and from these data, he deduced the relative contributions—or <u>*importance*</u>—of different frequency regions to the articulation score (the most important region is between 1-2 kHz), called the articulation distribution, *D*. He also determined the contribution of speech energy as a function of frequency of the communication system being evaluated, *W*, derived from the system's frequency response. Crandall arrived at the predicted articulation score by integrating (or summing) the area under the $D \times W$ curve.

Crandall's 1916 report introduced three fundamental concepts that were later used by Fletcher to calculate the *AI*: the articulation distribution (or *importance function*) and the energy contribution, both just described, and the <u>*additive*</u> <u>*property*</u>. The additive property is the assumption that the audible frequency spectrum can be divided into separate frequency bands, each band makes an independent contribution to the articulation score, and the separate contributions can be added together to yield a performance indicator. For example, if a 1-kHz high-pass filtered condition yields a 75% correct articulation score and a 1-kHz low-pass condition yields 25%, then the additive property implies that both bands presented together will yield 100%. Crandall compared his predicted articulation scores to actual scores for two communication systems and found reasonably good agreement.

<u>*FLETCHER (1921).*</u> To check Crandall's method, Fletcher collected a fresh set of articulation scores with newly-designed "ideal" filters and better control of speech intensity levels. Contrary to Crandall's conclusions, Fletcher found that Crandall's importance function *D* was not accurate and that articulation scores did not demonstrate the additive property. Fletcher recognized that the solution was to create an index that had both the additive property and a known relationship with intelligibility. That is, instead of calculating the articulation score directly, Fletcher proposed to calculate an index and then infer the corresponding articulation score from the index. The index would be calculated using the physical transmission characteristics of the communication system.

In his 1921 report, Fletcher's presented a new metric called the quality index, *Q*, that he later called the articulation index, *A* (or *AI*). He defined the *AI* as a number ranging from 0 to 1 whose value increases monotonically with increases in articulation score. Fletcher began by determining the *AI*'s relationship with articulation scores from the results of filtered speech experiments. For example, he noticed that 1.930-kHz high-pass and 1.930-kHz low-pass filtered conditions each yielded an articulation score of 70% correct. He concluded that the two conditions must have the same *AI* because they give the same articulation score. When combined together, the two conditions covered the full range of the frequency spectrum important for speech reception, implying $AI = 1.00$. Assuming the additive property, Fletcher deduced that the *AI* for each of the filtered conditions must be 0.50. This gave him one point on the curved line relating *AI*'s to articulation scores. Following this process of deduction, he was able to define the entire function.

To calculate the *AI*, Fletcher found it necessary to consider the speech intensity level, noise levels, and other distortions, in addition to a system's frequency response that was used by Crandall. Fletcher included quantities from basic hearing research such as, for example, detection thresholds, auditory filter bandwidths, and loudness growth functions. He also used a database of articulation scores that were collected using an ideal transmission system that had been constructed at Bell Labs specifically for comparing proposed communication systems to optimal transmission.

Fletcher's general *AI* equation was: $AI = V \times E \times F \times C$. The equation implies that there are four percepts or dimensions that contribute independently to speech intelligibility. Each of the four factors takes on values in the range from 0 to 1, as does the *AI*. Each factor correlates with an acoustical parameter of the communication system being studied. *V*, the volume factor, indicates the audibility or loudness of the speech contained in the audio signal. *V* is associated with the speech-to-noise ratio in the audio signal. *V* increases when more gain is applied to the speech. *E*, the ear sensitivity factor, is an intensity limit such that when *E* is less than 1, speech is so intense that it decreases speech intelligibility. *F*, the frequency distortion factor, quantifies spectral balance or the flatness of the frequency response. *F* equals 1.00 when the frequency response is flat, and is less than 1.00 by an amount that depends on how much it differs from flat. [The *F* factor contains the concept of the importance function originally proposed by Crandall.] *C* is the carbon distortion factor. It was intended to account for the types of nonlinear distortions introduced by carbon transmitters that were in use at the time. *C* was later renamed *H* and was extended to include ear-generated intermodulation distortion, reverberation, frequency compression, frequency shifting, peak-clipping, and amplitude compression.

The significance of the form of the general equation is that it provides diagnostic capabilities and directs improvements. For example, if *E* is found to be less than 1.00 for a particular speech communication system—meaning speech is too loud---then decreasing the gain may increase speech intelligibility. Similarly, smoothing the frequency response may increase the *AI* by increasing the *F* factor.

FLETCHER AND GALT (1950). The final, finely-tuned version of the *AI* calculation was published in 1950, nearly three decades after the first version. It was originally issued as an internal memorandum in 1947. During the years between 1921 and 1947, the Bell Labs group repeatedly evaluated and revised the *AI* calculation based on their experimental work in acoustics and audition, and tested the revisions using their growing speech sound recognition database. The 1950 *AI* is the culmination of all of this work, and was published along with an extensive validation study that demonstrates its accuracy in predicting articulation scores for a broad range of conditions. Although the 1950 calculation is a substantially different from the 1921 version, the general equation remains the same: $AI = V \times E \times F \times H$.

FRENCH AND STEINBERG (1947), Beranek (1947), ANSI S3.5-1969. Articulation index research was discontinued when Fletcher retired from Bell Labs. Consequently the 1950 *AI* was never used in practice. The first American National Standards Institute version, ANSI S3.5-1969, was actually derived from a simpler *AI* calculation provided by Bell Labs to the Harvard's Psycho-acoustic and Electro-acoustic Laboratories in 1942 (French, 1942) to help WWII communications research. Two versions of the wartime *AI* were published in research journals after the war: French and Steinberg (1947) (identical to French, 1942) and Beranek (1947). Beranek, who led Harvard's Electro-acoustic Laboratory, modified French (1942) to accommodate military applications. The two 1947 papers were the first to reveal the *AI* concept publicly since all prior versions were confidential internal reports. It is often incorrectly assumed that the French and Steinberg version is the definitive *AI* because it has the earliest publication date. French and Steinberg indicate that the *AI* was derived from earlier work at Bell Labs, but Crandall 1916 and Fletcher 1921 were not specifically cited probably because they had not been published.

Evaluations of the French and Steinberg ensued, and these demonstrated certain limitations that also persist in the ANSI implementations. Most importantly, evaluations have found inaccurate predictions of speech intelligibility test scores for some types of distortion, particularly, a discrepancy between noise and filtered-speech conditions. This was attributed to faulty handing of conditions containing noise. Fletcher was aware of this limitation, knew his version could handle noise, and this was one reason he gave for wanting to publish his 1950 version. Even though it is superior, Fletcher and Galt's 1950 *AI* continues to be overlooked presumably because it is computationally intensive and it contains concepts that require familiarity with the speech and hearing sciences.

Kryter led the effort to create an American National Standards Institute version. He considered Fletcher's calculation, but chose to modify Beranek's 1947 version for ANSI S3.5-1969 because it was a simpler calculation. Simplicity was important at that time because computers were not generally available to likely *AI* users. Kryter published a validation study demonstrating that his ANSI version worked reasonably well for filtered speech-in-noise conditions.

Speech Intelligibility Index (ANSI S3.5-1997). ANSI S3.5-1997 renamed the *AI* the "speech intelligibility index" (SII) and this change appropriately reflects significant departures from ANSI S3.5-1969. The SII calculation has the same general form of ANSI S3.5-1969, but introduces the *type* of speech materials (e.g. nonsense syllables, words, sentences, or connected discourse) as a variable in the calculation. This effectively voids one of Bell Labs' major accomplishments: the creation of a metric that is immune to speech materials differences. That is to say, it disregards an early goal of the *AI* which was to characterize a device "depending only upon its physical characteristics and the noise conditions at the listener's ear." Another significant departure is that the SII encourages users to customize the calculation by supplying their own speech spectra and weighting functions, values that were constants in all previous *AI* versions. This diminishes the SII's value as a standard by preventing comparisons across laboratories. It also complicates attempts to validate the SII. To date, no systematic validation of the SII has been conducted.

FLETCHER (1952). Fletcher and his group had a long-standing interest in measuring, classifying, and remediating hearing loss. Bell Labs assembled hearing aids, built the first commercial audiometer, and regularly conducted evaluations of hearing-impaired persons who contacted Bell Labs for help. Fletcher wrote scientific papers and trade journal articles, and lectured to medical groups concerning hearing loss and hearing aid selection and fitting. Fletcher was the first to apply the *AI* to cases of hearing impairment and hearing aid fitting, but his effort was not completely successful. It seems that reducing the calculation to make it practical for clinical application may have caused the failure.

RANKOVIC (1995, 1998, 2002). Both Jont Allen and Rankovic recognized that Fletcher's work was valuable and should be revived. Whereas Allen focuses on theoretical underpinnings and their applications to automatic speech recognition, Rankovic's work relates to audiology applications. She published three studies whose results show that Fletcher's 1950 *AI* predicts accurately for cases of hearing impairment. Test conditions included amplified speech in quiet and in noise, with hearing-impaired listeners having a variety of hearing loss configurations including those thought to have "dead" cochlear regions. Specific findings were:

> *Fletcher's AI is superior to ANSI for normal-hearing and hearing-impaired listeners*. Rankovic (1995) compared the Fletcher and Galt (1950) calculation to the ANSI S3.5-1969 calculation. She found that Fletcher's calculation resolved discrepancies between speech-in-noise and filtered-speech conditions for normal-hearing listeners. She also found that discrepancies between normal-hearing and hearing-impaired listener groups were resolved when Fletcher's calculation was used.
>
> *Fletcher's AI is superior to ANSI for amplified speech in intense band of background noise.* Rankovic (1998) collected masking patterns of octave-band noises from hearing-impaired listeners presented with amplified speech in a background of the same noises. She found that Fletcher's *AI* accounted for complicated interactions between the audiogram, frequency response, gain, and masking caused by the noise.
>
> *Fletcher's AI accounts for so-called "dead" cochlear regions.* Rankovic (2002) examined the speculative claim of Brian Moore and colleagues (2001) that the articulation scores of hearing-impaired listeners described as having cochlear "dead regions" could not be predicted by the *AI*. Dead regions of the cochlea are thought to have no functioning inner hair cells. She calculated *AI*'s for their data and found that the *AI* predicted the scores, with no special consideration given to the fact that regions were "dead."

Interactive *AI™.* **Articulation Inc** has constructed the Interactive *AI*™ to make the original Bell Labs *AI* accessible via an easy-to-use graphical interface. **Articulation Inc** is committed to calling attention to this established, science-based model of audition and speech intelligibility. We continue to conduct validation studies, create extensions, and improve the graphical user interface to accommodate user needs and broader application of this important and unique tool.