JASA, vol(19) No(1), Jan 1947

Factors Governing the Intelligibility of Speech Sounds

N. R. French and J. C. Steinberg

1. Introduction

- Goal: Determine a quantitative relationship between the fundamental characteristics of speech and hearing, and the capability of the ear in recognizing the speech sounds.
- Problem: The ability of a listener in recognizing a sound depends on a lot of factors, such as intensity of speech, intensity of noise, echo, reverbration, hearing, etc. Also there is the variations of response and hearing acuity.
- Strategy: Ignore the minor factors, consider only the major factors, such as the basic characteristic of speech and hearing; the electrical and acoustical characteristics of the communication system.

2. Basic Concept

The formulation of AI is based on three basic observations.

- Different frequency bands are independent from each other in contributing to the intelligibility of speech sound.
- Speech spectral is generally independent of speech intensity. In other words the frequency range can be divided into multiple bands without consideration of the speech intensity.
- The articulation index of any narrow band is proportional to the fraction of syllables that can be heard in that frequency band. Those cannot be heard surely will not take any effect.

2. Basic Concepts

- Definition: $A = \sum W \cdot \Delta Am$, $A \in [0, 1]$.
- Different frequency bands are independent from each other in contributing to the intelligibility of speech sound.
- For convenience in making calculation, the 20 frequency bands are chosen such that $\Delta Am = 0.05$.
- W is the fraction of 1/8th second intervals in which the speech intensity in the particular band is of sufficient intensity to be heard, in other words, the effective sensation level $E > 0$dB.
- E depends on three kinds of masking: 1) forward masking; 2) upward masking; 3) inband masking.

3.1 Long term average speech spectrum

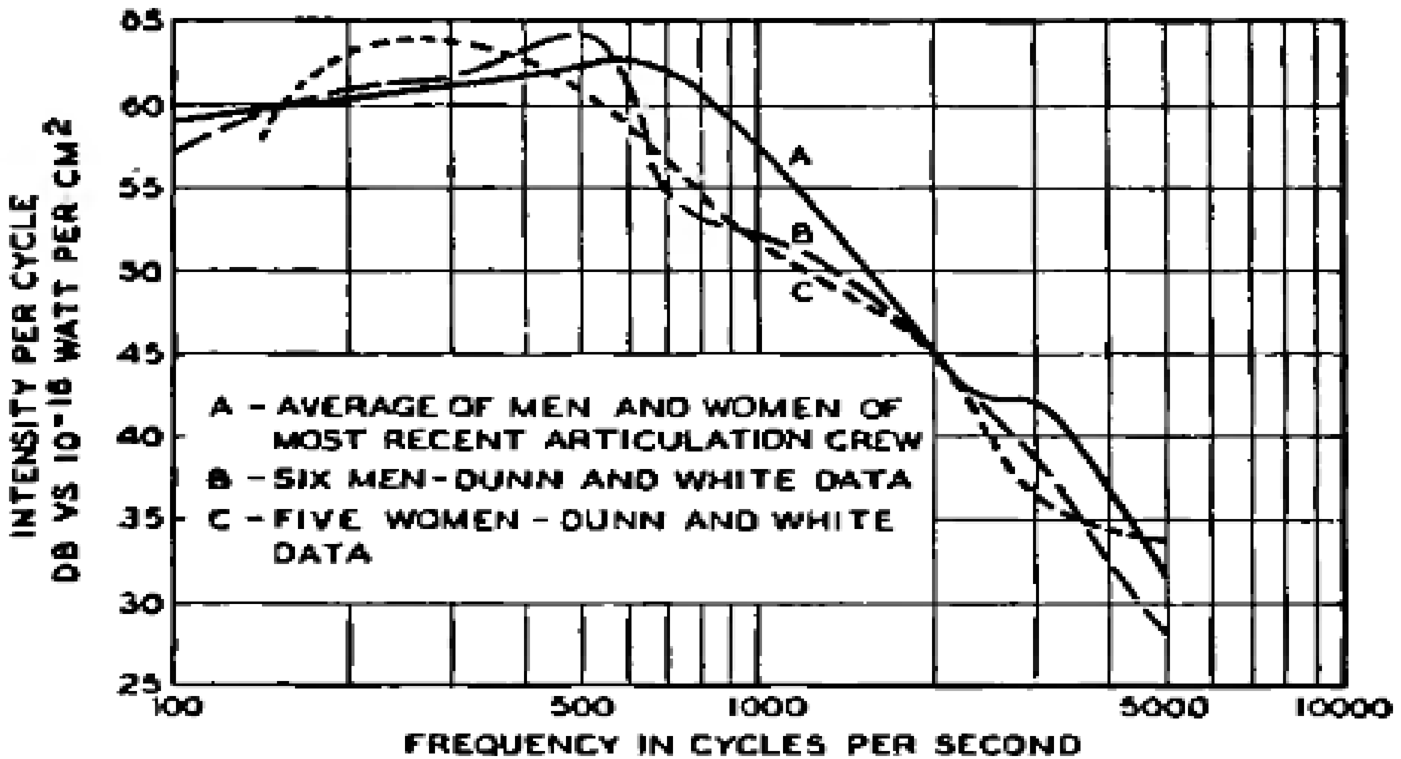

FIG. 1. Comparison of speech spectra at two inches from lips.

Calculation of AI is based on long-term average speech spectrum!

3.2 Level Distribution of Speech

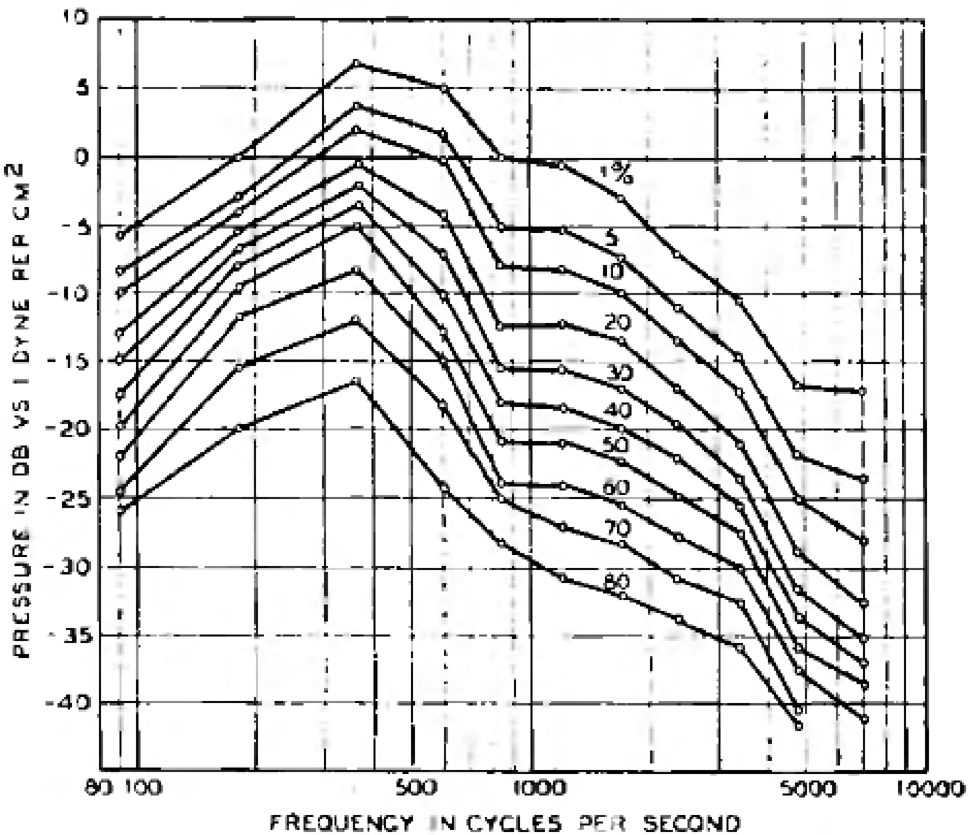

FIG. 3. R.m.s. pressure, during one-eighth second intervals, of speech at 30 cm from lips. Dunn and White composite data for six men (reference 6). Each curve shows the pressure exceeded in the indicated percentage of intervals.

- Speech spectrum is independent of the intensity. Spectrum of 1% peaks and 80% peaks have almost the same shape.
- As a consequence, the frequency range can be divided into 20 bands of equal contribution without considering the level of the stimuli.

3.4 Masking

- There are three kinds of masking: 1) forward masking; 2) upward masking; 3) inband masking.
- The total masking of sound is given by $M = (B (+) X) - X + m$, where $B(+)X \equiv 10 \log10(10 B/10+10X/10)$
- B is the power spectral density of noise in a certain frequency band, including in-band masking and upward masking.
- X is the 0db, or say just noticeable threshold of spectral power density characteristic of human ear;
- m is a correction factor used when the effective level of noise is above 50db, it accounts for the forward masking.

4. Response Characteristics

- Purpose: get an accurate prediction of Bs, the long average speech spectrum received by the ear.
- Question: How to compensate for the characteristics of communication systems, including air path?
- Definition: a telephone system has an ***orthotelephonic response of zero db*** at all frequencies when it can be replaced by a one-meter air path, between talker and listener, without changing the loudness of the received speech at any frequency.
- If a communication system has an orthotelephonic response of R dB, Bs = Bs' + R
- In addition, if the actual speech level V is different from 90db, correction can be made by Bs = Bs' + (V - 90) + R

5. Articulation Index

- Motivation: 1) syllable articulation is dependent on the skill and experience of the listener; 2) the articulation score is not additive across frequency bands.
- Fletcher-Stewart 1921 product rule: $e = e1*e2*e3*...e20$
- Apply log on both sides ⇒ $\log(1-s)=\log(1-s1)+\log(1-s2)+...$
- Define: $A = \log(1-s)$, $\Delta A = \log(1-sk)$, we have $A = \sum \Delta A$, where $\Delta A = W \cdot \Delta Am$ is the information carried by any narrow band.
- Thus $A = \sum W \cdot \Delta Am$. The determination of AI involves two steps: 1) determination of frequency increment that gives equal values of ΔAm through out frequency range; 2) determination of relationship between W and the levels of speech and noise.

Determination of Frequency Bands

- Method: low-pass / high-pass filtering at multiple snr conditions to measure the frequency importance function.
- Problem: the articulation score S is not additive, it can not be used directly to measure the importance of any frequency band.
- Step 1: convert the articulation score S into AI.
- Step 2: Divide the frequency range into 20 bands based on AI.

Determination of Frequency Bands

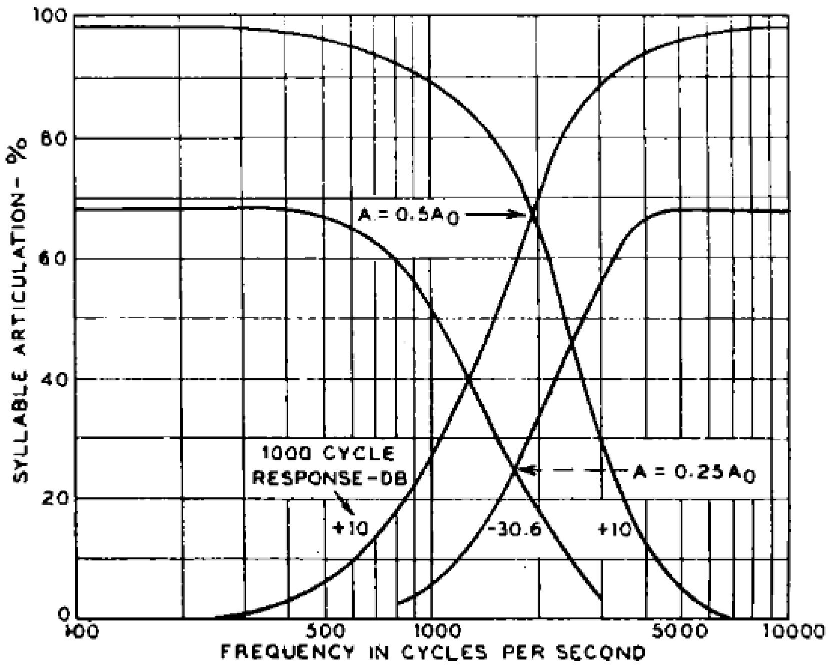

FIG. 12. Syllable articulation *versus* cut-off frequency of high pass and low pass filters at two different settings of test circuit. A_0 is articulation index of test circuit at its optimum setting.

- Assume: Smax (0.98) ⇔ A0, the point where LPF curve and HPF curve meet is 0.5A0. i.e. 0.67 ⇔ 0.5A0
- Add noise, make Smax' = 0.5A0, then 0.25 ⇔ 0.25A0
- S(LPF=3200Hz) ⇔ 0.25A0
- S(HPF=3200Hz) ⇔ 0.75A0
- 0.90 ⇔ 0.75A0

Determination of Frequency Bands

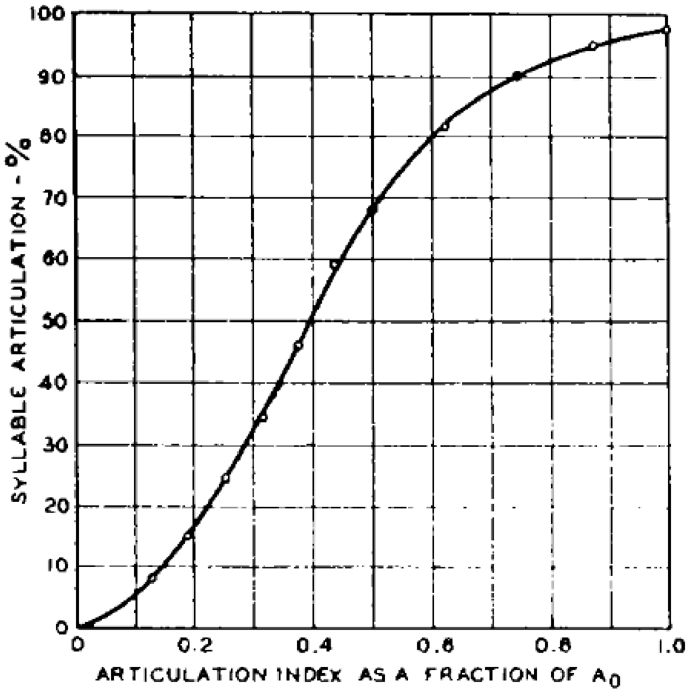

FIG. 13. Relation between syllable articulation and articulation index. The latter is expressed as a fraction of the articulation index (A_0) of the test circuit at its optimum setting.

- Following the procedure, the relation between syllable articulation S and AI is plotted in Fig. 13.

Determination of Frequency Bands

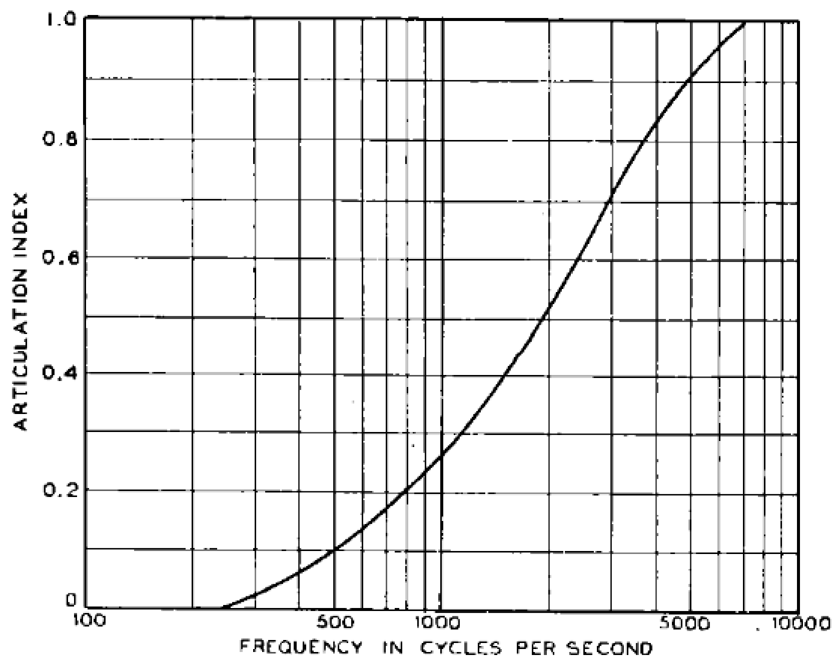

FIG. 16. Articulation index *versus* cut-off frequency. All bands are at their optimum levels. Curve is based on about equal numbers of men's and women's voices.

- Re-plot the S vs cut-off frequency under clean condition and replace S with AI. Fig. 16 depicts the AI vs cut-off frequency.
- The frequency bands of equal ΔAm can be determined directly from Fig. 16.

Determination of Frequency Bands

TABLE III. Frequency bands making equal (5 percent) contributions to articulation index when all bands are at their optimum levels. Composite data for men's and women's voices.

Band	Frequency limits cycles	Band	Frequency limits cycles
1	250–375	11	1930–2140
2	375–505	12	2140–2355
3	505–645	13	2355–2600
4	645–795	14	2600–2900
5	795–955	15	2900–3255
6	955–1130	16	3255–3680
7	1130–1315	17	3680–4200
8	1315–1515	18	4200–4860
9	1515–1720	19	4860–5720
10	1720–1930	20	5720–7000