

Review of Objective Speech Quality Measures for Voiceband Coders

Maurizio Copperi

1. INTRODUCTION

The purpose of this paper is to review the most important published studies in the area of objective measures, useful for the quality prediction of speech coders. Admittedly, "quality" is an extremely elusive concept that involves something more than intelligibility and takes into account the phenomenon of individual differences in taste.

Indeed, a primary merit of subjective measures is that they provide a score which incorporates all aspects of the human speech perception process. In fact, listeners do not use only the acoustic cues but actively exploit their knowledge of the language, the syntactic and semantic contexts and even talker related information to succeed in the perception task. However, either many subjects or trained crews must be used in the experiments, in order to remove the large component of variance due to the relative tolerance of each subject for distortions and noise. Unfortunately, subjective scores are not related to physical signal characteristics and give little insight into how a speech codec can be improved. Another drawback is that subjective measures are susceptible to errors of both subjects and administrators. Further, it is difficult to compare subjective results obtained in different times and places.

In the recent years, new digital speech processing techniques have been developed and incorporated in voiceband codecs. Because of the large number of factors determining the output speech quality, it is practically impossible, as previously noted, to optimize the coding system by resorting to subjective judgments, which are generally costly and time consuming.

Therefore, various objective measures have been tested and compared in order to establish their degree of correlation with subjective scores. Objective measures are inexpensive to administer and quite reliable. In addition, it is

SIP Headquarters, Research and Development Division, Via S. Dalmazzo 15, 10122 Torino (Italy)

easy improving the coder performance by directly minimizing the distortion as defined by the objective measure itself.

So far, a clear-cut answer to the problem of the quality evaluation of very different coders or speech distortions through a single universal objective measure is not yet available. In fact, an optimal quality estimator should provide not only a good prediction of the performance of a given coder, but also a correct ranking among various coders. However, the results presented in the literature, and summed up in the following, show that a few objective measures look promising for certain classes of speech digitizers.

2. OBJECTIVE MEASURES

This section deals with the description of a number of computable objective measures that are usually employed as tools in the evaluation task. The first three measures are defined in the time domain and represent a simple way to characterize, in a single number, the performance of a codec under test. The other measures are defined in the frequency domain and permit a more sophisticated approach to the issue of gauging speech quality. Moreover, they are insensitive to short delays between input and output signals or to phase distortion.

2.1. Long-term SNR

A widely used measure of performance, easy to compute and well understood, is the conventional signal-to-noise ratio (SNR) defined as

$$\text{SNR} = 10 \text{ Log } \left\{ \frac{\sum_n x^2(n)}{\sum_n [y(n) - x(n)]^2} \right\} \quad (1)$$

where $x(n)$ and $y(n)$ are the input and output signals, respectively. Since the summations are taken over the entire speech utterance, Eq. (1) is called long-term SNR.

Several experiments have shown that this measure is poorly correlated with subjective quality, nevertheless is sometimes used during the design and "tuning" of waveform coders.

2.2. Segmental SNR (SNRseg)

An improvement to the previous measure, suggested by P. Noll [1], averages SNR values over short (15-30ms) segments and therefore assigns equal weight to loud and soft parts of the utterance. This measure can capture individual preferences for a given coder, but still fails to predict a correct ranking between different coders. The main shortcoming of this measure arises when the variation of individual SNR values around the average is large.

2.3. Gain-compensated SNRseg

In this measure, variations of the output speech level with respect to the input signal are compensated before taking the SNR, segment by segment. This procedure is supported by the fact that small amplitude variations introduced by the coder can impair the SNR measurements, while having a negligible impact on the subjective quality.

2.4. Frequency weighted segmental SNR

Many experiments have shown that the perceptual quality of a coder depends, among other things, on the frequency distribution of this quantizing noise relative to the speech spectrum. In fact, it is well known that the auditory mechanism relies upon a short-term spectral analysis of the incoming signal, exploiting this spectral information as a frequency warped "place spectrum" translated on the basilar membrane of the ear [7]. It turns out that a speech signal is judged to be of "good quality" when each location on the basilar membrane (or equivalently each "critical band") is excited by a signal with a sufficiently high SNR.

According to the classical articulation model, the speech band, ranging from 200 up to 6100 Hz, is divided into 20 nonuniform subbands experimentally derived. Each subband is assumed to contribute, independently of the others and under optimum conditions, an equal 5% to the so-called articulation index (AI)

$$AI = \sum_j A(j) = \sum_j 0.05(SNR(j)/TH), \quad j = 1, \dots, 20 \quad (2)$$

where the peak signal to rms noise ratio in band j , $SNR(j)$, is clipped to a maximum value of TH dB (e. g. $TH = 30$ dB), so that AI cannot be greater than 1.

Eq. (2) can be transformed to an integral in frequency [7], giving

$$AI = \int_{f_1}^{f_2} F(f) SNR(f) df, \quad f_1 = 200, \quad f_2 = 6100 \quad (3)$$

where $F(f)$ (the Jacobian of the transformation) is a frequency weighting that falls down with an approximate slope of 20 dB from 200 to 6100 Hz. Therefore, the AI measure can be considered as a frequency weighted SNR measure.

An important feature of this measure is that the width of each band increases with the center frequency, in order to carry the same amount of contribution. This is in tune with the fact that the short-term speech spectrum tends to be flatter (whiter) in high frequency subbands, which leads to a decreasing amount of information (entropy). Moreover, the weighting behavior is in agreement with the optimal SNR distribution, as a function of frequency, for subband coders.

Considering the time varying nature of the short-time speech spectrum, a static weighting function does not perform adequately well, while a dynamic one has the potential to yield a better measure of quality. A general form assumed by this refined measure is

$$SNRF = \sum_i \int_{f_1}^{f_2} G[S(f)]_i F(f) SNR(f)_i df \quad (4)$$

where i refers to the speech segment index, f is frequency, $S(f)$ is the short-time spectrum of the input speech, $G[S(f)]_i$ is a dynamic frequency weighting which is related to the speech production mechanism, $F(f)$ is a static frequency weighting derived from psycho-acoustic properties of hearing, and $SNR(f)$ is the short-time SNR at frequency f (i.e. the coder performance).

From Eq. (4), we can define a number of objective measures, which differ in the dynamic weighting function G and in the SNR computation [8].

The first form is based on the assumption that $G[S(f)] = 1$, and $SNR(f)$ is computed in dB (log scale). So we have

$$SNRF_1 = \langle \sum_j [10 \text{Log}(\mu_{xj} / \mu_{nj})] \int B \rangle, \quad j = 1, \dots, B \quad (5)$$

where μ_{xj} is the power of the reference signal $x(n)$ in the articulation band j , for a short segment (e.g. 30ms), and μ_{nj} is the power of the corresponding noise signal $y(n) - x(n)$. The notation $\langle \rangle$ denotes the average over all the segments in the speech utterance.

The second form is characterized by a weighting that approximates the subjective loudness

$$G[S(f_j)] = L(f_j) = \int_j |S_y(f)|^{-0.5} df \quad (6)$$

where $S_y(f)$ is the spectrum of the coded speech, and f_j is the center frequency of band j . The resulting measure is

$$SNRF_2 = \langle \left\{ \sum_j L(f_j) [10\text{Log}(\mu_{xj}/\mu_{nj})] \right\} / \sum_j L(f_j) \rangle, \quad j = 1, \dots, B \quad (7)$$

The third form exploits a normalized log spectral weighting

$$G[S(f_j)] = P(f_j) = [40 + 10\text{Log}(\mu_{yj}/\mu_y)] / 40 \quad (8)$$

where μ_{yj} is the output signal power in band j , and μ_y is the total power in the output signal (for the i -th segment). Thus $P(f_j)$ is 1 if the energy in a band j is equal to the total energy, and is set to 0 if it is 40 dB (or more) below the total output energy. Therefore, the log spectral weighted measure is

$$SNRF_3 = \langle \sum_j P(f_j) [10\text{Log}(\mu_{xj}/\mu_{nj})] / B \rangle, \quad j = 1, \dots, B \quad (9)$$

2.5 Cepstral distance measures

While the preceding SNRF's are based on the entire spectrum, computed through FFT, other distance measures are based on transformations that retain only the smoothed spectral behavior of the speech signal.

We will consider first the unweighted Euclidean distance based upon cepstral coefficients, which will be referred to as the cepstral distance measure CDM.

Consider two all-pole spectral models $G/A(z)$ and $G'/A'(z)$. The error or difference between these models on a log magnitude versus frequency scale is defined as [32]

$$V(\vartheta) = \ln [G^2 / |A(e^{j\vartheta})|^2] - \ln [G'^2 / |A'(e^{j\vartheta})|^2] \quad (10)$$

where ϑ a normalized frequency or angle in the z plane, with π representing the half-sampling frequency. A logical choice for a distance measure between spectral models is the set of L_p norms defined as

$$(d_p)^p = \int_{-\pi}^{\pi} |V(\vartheta)|^p \frac{d\vartheta}{2\pi} \quad (11)$$

The rms log spectral measure is defined for $p = 2$. These L_p measures can be related to decibel variations in the log spectral domain through the multiplicative factor $10/\ln(10) = 4.34\dots$

In order to reduce the computational load required to estimate $V(\vartheta)$ as a summation, we can resort to other efficient methods based on linear prediction analysis. To this effect, if $A(z)$ is an M th order polynomial in z^{-1} with all of its roots within the unit circle, and $A(\infty) = 1$, then a Taylor series expansion gives

$$\ln [A(z)] = -\sum_k c_k z^{-k}, \quad k = 1, \dots, \infty \quad (12)$$

where $\{c_k\}$ are the cepstral coefficients. It follows [32] that the Fourier series expansion for the model log spectrum is

$$\ln [G^2 / |A(e^{j\theta})|^2] = \sum_k c_k e^{-jk\theta}, \quad k = -\infty, \dots, \infty \quad (13)$$

where $c_0 = \ln [G^2]$ and $c_{-k} = c_k$.

An application of the Parseval's theorem to the L_2 distance measure gives

$$(d_2)^2 = \sum_{k=-\infty}^{\infty} (c_k - c'_k)^2 = (c_0 - c'_0)^2 + 2 \sum_{k=1}^{\infty} (c_k - c'_k)^2 \quad (14)$$

All the spectral shape information lies in the coefficients c_1, c_2, \dots, c_M , since they uniquely describe the filter coefficients of $A(z)$. Thus, we can take a truncated series to define a cepstral measure $u(L)$, for L greater than or equal to M , as

$$[u(L)]^2 = (c_0 - c'_0)^2 + 2 \sum_{k=1}^L (c_k - c'_k)^2 \quad (15)$$

The cepstral distance measure in dB is defined as:

$$\text{CDM} = (10/\ln 10) [2 \sum_{k=1}^L (c_k - c'_k)^2]^{0.5} \quad (16)$$

In conclusion, Eq. (16) can be readily computed by means of linear prediction analysis, to evaluate the models $A(z)$ and $A'(z)$, and well known transformations between model coefficients a_j and cepstral coefficients c_k .

It is also possible to define the quefrency weighted cepstral distance or Root Power Sums [33] as

$$\text{DqW} = \sum_{k=1}^L k^2 (c_k - c'_k)^2 \quad (17)$$

The most important feature of the weighting is that it de-weights the lower order cepstral coefficients, rather than weighting the higher order ones, in a data independent manner.

Finally, it is worth mentioning the use of transitional spectral variations into a specific cepstral measure. This dynamic feature of the spectral space has mainly been used in speech recognizers [34]. An application to speech compression algorithms is presented in [35].

The spectral variation in time is represented by the time derivative of the log-spectrum or, recalling Eq. (13), by the time derivative of the sampled time series $c_k(t)$, that usually does not have an analytic form. Since the 1st order finite difference is in general noisy, the derivative can be approximated by an orthogonal polynomial fit on each cepstral trajectory over a fixed number of frames (window).

The 1st order coefficient, or spectral slope in time, of the orthogonal polynomial has the form

$$\Delta c_k(t) = [\sum_n n h_n c_k(t+n)] / [\sum_n h_n n^2], \quad n = -N, \dots, N \quad (18)$$

where h_n is the window of length $2N + 1$. A weighted Euclidean distance between two given transitional spectra is defined as

$$Dw = \sum_k w_k (\Delta c_k - \Delta c'_k)^2 \quad (19)$$

Dynamic spectral features play an important role in speech perception, as demonstrated in a perceptual experiment by S. Furui [36].

2.6 Other parametric and spectral measures

In addition to the previous objective measures, other parametric and spectral measures have been suggested for the evaluation of speech coders [25]. Of particular interest are those based on linear prediction analysis, such as: log area ratio (LAR) measure, reflection coefficient (RFC) measure, feedback (or predictor) coefficient (FBC) measure, log likelihood ratio (LLR) measure,

linear spectral distance (LSD) and frequency-variant spectral distance (FVSD).

The measures LAR, RFC and FBC are computed by performing the linear prediction analysis over input and output speech frames, and then evaluating the L_1 norm between the corresponding input and output parameters.

For LLR, the likelihood ratio between input and output LPC parameters is raised to the power 0.25.

For LSD, the LPC all-pole model spectra for input and output speech are normalized to have the same geometric mean, then the L_2 norm is taken.

The measure FVSD is also computed using the input and output LPC model spectra. Each spectrum is divided into typically six bands, which are separately normalized so that the average spectral amplitude over each band is unity. A weighted L_4 norm is taken between the normalized subband spectra, the weighting function being the input LPC spectrum. Finally a linear combination of the six norms is formed. The constants required for the linear combination can be obtained via linear regression analysis with subjective scores.

The short-time banded SNR (STB-SNR) is a generalization of the segmental SNR. The input speech and the noise (difference between the input and the output signals) are filtered into six bands and for each band the SNRseg is computed. A linear combination of the six SNRseg values is then formed. The final objective score is obtained, as usual, by time averaging over frames.

Another interesting measure, proposed in [26], incorporates an explicit parametric model of speech perception, and is based on the perturbations exhibited by the spectral peaks of the output signal. Speech formants are computed on the original and distorted signals via the Line Spectrum Pair transformation of the LPC polynomial. Nine different features (e. g. energy, differences in location and in bandwidth, movement, etc.) are determined by comparing the spectral peaks, classified as lost, false and distorted. The objective measure is obtained as a combination of 13 individual features, each computed on at most four spectral peaks.

3. EXPERIMENTAL RESULTS

In this section, we report briefly the most important results obtained by various researchers. In particular, each subsection deals with the work performed in a specific laboratory.

3.1 BBN (U.S.A.) [2.5]

Selected coders:

- Adaptive Predictive Coding with Noise Shaping (APS-NS), bit rate of 16 kbit/s,
- Continuously Variable Slope Delta modulation (CVSD) at 16 kbit/s,
- APC with Segmented Quantizer (APC-SQ) at 9.6 kbit/s,
- Base Band Coder (BBC) at 9.6 kbit/s,
- LPC vocoder with 10th order prediction filter (LPC-10) at 2.4 kbit/s.

Subjective measure:

- Diagnostic Acceptability Measure (DAM).

Objective measures:

- log area ratio,
- reflection coefficients,
- predictor coefficients,
- log likelihood,
- linear spectral distance,
- frequency-variant spectral distance,
- short-time banded SNR.

Results: FSVD has the best correlation (0.997) for the waveform coders, followed by STB-Snr (0.994).

3.2 Bell Laboratories (U.S.A.) [6-11]

Selected coders:

- Adaptive Differential PCM (ADPCM) with Jayant adaptive quantizer and fixed predictor, bit rates of 9.6, 16, and 24 kbit/s,
- ADPCM with forward adaptive quantizer and forward adaptive 8th order predictor, bit rates of 9.6, 18, and 26 kbit/s,
- Sub Band Coder (SBC) at bit rates of 9.6, 16, and 24 kbit/s,
- Adaptive Transform Coder (ATC) at bit rates of 9.6, 16, and 24 kbit/s.

Subjective measure:

- Absolute category rating with Mean Opinion Score (MOS).

Objective measures:

- SNR, SNRseg,
- frequency weighted SNRseg,
- log likelihood,
- articulatory bandwidth.

Results: A combined measure, based on a modified log likelihood and the percent articulatory bandwidth, predicts well the individual preferences in a given coder as well as the inter-relationships between coders.

3.3 Bell Northern Research (Canada) [12-14]

Selected coders:

- μ -law PCM ($\mu = 255$), from 4 up to 8 bit/sample,
- Adaptive Delta Modulation (ADM) at bit rates of 16 and 32 kbit/s,
- ADPCM, with either adaptive or fixed 3rd order predictor, at 32 kbit/s,
- APC at 16 kbit/s,

- Residual Excited Linear Predictive coder (RELP) at 4.8 kbit/s,
- LPC vocoder at 2.4 kbit/s.

Subjective measures:

- preference test (ABAB, BABA sequences),
- MOS.

Objective measures:

- SNR, granular SNR, overload SNR, subjective SNR,
- coherence function.

Results: The subjective SNR, defined as the SNR of that reference signal which, on the average, is equally preferred to the test signal by a group of listeners, can be used only for waveform coders of good quality. The correlation obtained using the coherence function over PCM, ADPCM and APC is 0.96.

3.4 CSELT (Italy) [15-17]

Selected coders:

- ADPCM, with fixed predictor, at 24 and 32 kbit/s,
- APC-NS based on FFT.

Subjective measure:

- MOS on a 9 point scale.

Particular condition:

- ADPCM with channel bit error rate (BER) equal to 1/256 and 1/32.

Objective measures:

- SNR, SNRseg,
- frequency weighted SNRseg,

- log likelihood,
- several linear combinations of these measures.

Results: The best prediction scores are achieved by a gain-compensation SNRseg, and also by a spectral signal-to-distortion ratio. A linear combination of these measures has been used to predict ratings of APC-NS.

An attempt to estimate the quality of ADPCM coders through a measuring system, based upon artificial signals and identification procedures, is described in [17].

3.5 Georgia Institute of Technology (U.S.A.) [16-26]

Selected coders:

- Adaptive PCM (APCM), 6 conditions,
- ADPCM, 12 conditions,
- APC, 6 conditions,
- ADM, 6 conditions,
- CVSD, 6 conditions,
- SBC, 6 conditions,
- ATC, 6 conditions,
- LPC vocoder, 6 conditions,
- Voice Excited Vocoder (VEV), 12 conditions.

Particular conditions: more than 200 conditions with controlled distortions, such as low-pass, high-pass and band-pass filtering, center clipping, clipping, white noise, echo, etc., have been added to the database.

Subjective measure:

- DAM.

Objective measures:

- SNR, SNRseg,
- spectral distance,
- energy ratio,
- PARCOR coefficients,
- log area ratio,
- frequency-variant banded SNR,
- spectral peak distortion.

Results: A very good quality measure for waveforms coders and noise distortions is based on the frequency-variant banded SNR, which has a correlation factor of 0.93. The best simple measure is the spectral peak distortion, that gives the highest correlation factor (0.78) in comparison to any other simple measure previously tested on the Georgia Tech speech databases.

3.6 NTT (Japan) [27-30]

Selected coders:

- μ -law PCM ($\mu = 255$), from 4 up to 8 bit/sample, sampling frequency $F_s = 8$ kHz,
- ADPCM with a 2-pole 6-zero adaptive predictor and a backward adaptive dynamic locked quantizer (DLQ), from 2 up to 4 bit/sample, $F_s = 8$ kHz,
- ADPCM with a fixed 1st order predictor (coefficient = 0.98) and a backward adaptive quantizer (Jayant's algorithm), from 2 up to 4 bit/sample, $F_s = 8$ kHz,
- ATC with dynamic bit allocation among DCT coefficients, bit rates of 9.6, 12, 16 and 24 kbit/s,
- APC with Adaptive Bit allocation in time and frequency domains (APC-AB), split-band scheme with long term predictor (Honda-Itakura design), bit rates of 9.6, 12, 16 and 24 kbit/s.

Particular conditions:

- PCM with channel BER equal to 0.001,
- PCM overload.

Subjective measure:

- absolute category rating, giving MOS on a 5-point scale.

Objective measures:

- SNR, SNRseg,
- spectral distortion based on a 256-point FFT,
- cepstral distance,
- logarithmic spectral envelope distance, weighted by a cosh function,
- likelihood ratio (LHR),
- peak weighted LHR.

Results: Frequency domain measures have better correspondence to MOS than time domain ones; the cepstral distance has the best correlation (0.929), excluding BER and overload conditions.

3.7 Tokyo Institute of Technology (Japan) [31]

Selected coder:

- cepstral vocoder at a bit rate of 2.8 kbit/s, $F_s = 10$ kHz.

Subjective measure:

- preference score (pair comparison test).

Objective measure:

- Spectral distortion.

Results: There is a good correlation between preference scores and the overall spectral distortion caused by the frame period, cepstrum order and quantization noise.

4. CONCLUDING REMARKS

The large number of objective measures available so far in the literature represents an important result of the notable effort provided by many researchers in this field. However, the major problems continues to be the capability of a given objective measure to perform adequately well across a large sample of all distortions and all talkers. At the moment, based on experimental results and educated guess, we can say that this issue is twofold, since it involves both the statistical reliability of the objective measure and the correlation with subjective ratings. As far as the former point is concerned, it seems that the reliability is a very strong feature of many objective measures, while the latter issue is extremely subtle and deserves a careful examination.

Basically, the potential of an objective measure can be improved by tailoring its parameters on specific classes of distortions but this leads, unfortunately, to specialized quality estimators, losing generality and universality. On the other hand, a more general measure, devised to handle a wide range of speech distortions, will exhibit a comparatively lower performance, while requiring a huge subjective database to set the controlling parameters properly. In this light, it is clear that a substantial performance improvement could be provided by new measures designed according to an effective and advanced model of the speech perception process, rather than a signal fidelity criterion.

A few examples of measures conceived towards this goal have been reviewed in the preceding sections, but further enhancements are still needed to improve the quality prediction capability across a large set of different conditions and relevant distortions. This is a most basic step to be considered in the future research directions.

In conclusion, therefore, objective measures must be selected and used carefully, exploiting their usefulness for speech codec optimization and testing but, also, bearing in mind their current limitations.

REFERENCES

- [1] P. Noll: "Adaptive quantization in speech coding systems", Int. Zurich Seminar on Digital communication (IEEE), pp. B3.1-B3.6, 1976.
- [2] V.R. Viswanathan, W.H. Russel, A.W. Huggins: "Objective speech quality evaluation of mediumband and narrowband real-time speech coders", Proc. ICASSP, pp. 543-546, 1983.
- [3] V.R. Viswanathan, W. Russel, J. Makhoul: "Objective speech quality evaluation of narrowband LPC vocoders", Proc. ICASSP, pp. 591-594, 1978.
- [4] V.R. Viswanathan, J. Makhoul, W. Russel: "Towards perceptually consistent measures of spectral distance", Proc. ICASSP, pp. 485-488, 1976.
- [5] J. Makhoul, R. Viswanathan, W. Russel: "A framework for the objective evaluation of vocoder speech quality", Proc. ICASSP, pp. 103-106, 1976.
- [6] R. Crochieri et al.: "An interpretation of the log likelihood ratio as a measure of waveform coder performance", IEEE Trans. ASSP, vol. 28, N° 3, pp. 318-323, 1980.
- [7] R. Crochieri et al.: "A study of objective measures for speech waveform coders", Proc. Zurich Seminar, paper H1.1-H1.7, 1978.
- [8] J. Tribolet et al.: "A study of complexity and quality of speech waveform coders", Proc. ICASSP, pp. 586-590, 1978.
- [9] B. McDermott et al.: "Perceptual and objective evaluation of speech processed by adaptive differential PCM", BSTJ, vol. 57, N° 5, pp. 1597-1618, 1978.
- [10] D. Goodman et al.: "Objective and subjective performance of tandem connections of waveform coders with an LPC vocoder", BSTJ, vol. 58, N° 3, pp. 601-629, 1979.
- [11] L. Rabiner et al.: "Comparisons of system identification methods in the presence of high noise levels and bandlimited inputs", Proc. ICASSP, pp. 183-187, 1978.
- [12] P. Mermelstein: "Evaluation of a segmental SNR measure as an indicator of the quality of ADPCM coded speech", JASA, vol. 66, N° 6, pp. 1664-1667, 1979.

- [13] M. Nakatsui, P. Mermelstein: "Subjective speech-to-noise ratio as a measure of speech quality for digital waveform coders", JASA, vol. 72, N° 4, pp. 1136-1144, 1982.
- [14] "Objective evaluation of non-linear distortion effects on voice transmission quality", CCITT SG XII-1-E Annex 3, pp. 100-105.
- [15] C. Scagliola: "Evaluation of adaptive speech coders under noise channel conditions", BSTJ, vol. 58, N° 6, pp. 1369-1394, 1979.
- [16] N. Dal Degan, C. Scagliola: "Optimal noise shaping in adaptive predictive coding of speech", Proc. ICASSP, pp. 539-542, 1980.
- [17] R. Billi, C. Scagliola: "Artificial signals and identification methods to evaluate the quality of speech coders", IEEE Trans. Commun., vol. 30, N° 2, pp. 325-335, 1982.
- [18] T. Barnwell, A.M. Bush: "Statistical correlation between objective and subjective measures for speech quality", Proc. ICASSP, pp. 595-598, 1978.
- [19] T. Barnwell: "A comparison of parametrically different objective speech quality measures using correlation analysis with subjective quality results", Proc. ICASSP, pp. 710-713, 1980.
- [20] T. Barnwell: "Correlation analysis of subjective and objective measures for speech quality", Proc. ICASSP, pp. 706-709, 1980.
- [21] T. Barnwell, S. Quackenbush: "An analysis of objectively computable measures for speech quality testing", Proc. ICASSP, pp. 996-999, 1982.
- [22] S. Quackenbush, T. Barnwell: "Objective estimation of perceptually specific subjective qualities", Proc. 109th ASA Meeting, 1985.
- [23] P. Breitkopf, T. Barnwell: "Segmental preclassification for improved objective speech quality measures", Proc. ICASSP, pp. 1101-1104, 1981.
- [24] T. Barnwell: "Objective measures for speech quality testing", JASA, vol. 66, N° 6, pp. 1658-1663, 1979.
- [25] T. Barnwell, W. Voiers: "An analysis of objective measures for user acceptance of voice communications systems", Final Report, Contract DCA100-78-C-003, 1979.
- [26] H. Coetzee, T. Barnwell: "An LSP based speech quality measure", Proc. ICASSP, pp. 596-599, 1989.

- [27] N. Kitawaki, M. Honda, K. Itoh: "Speech-quality assessment methods for speech-coding systems", IEEE Commun. Magazine, vol. 22, Nº 10, pp. 26-33, Oct. 1984.
- [28] N. Kitawaki et al.: "Comparison of objective speech quality measures for voiceband codecs", Proc. ICASSP, pp. 1000-1003, May 1982.
- [29] K. Itoh, N. Kitawaki, K. Kakehi: "Objective quality measures for speech waveform coding systems", Review of ECL, vol. 32, Nº 2, pp. 220-228, 1984.
- [30] M. Sugiyama, K. Shikano: "Frequency weighted LPC spectral matching measures", Electronics and Commun. in Japan, vol. 65-A, Nº 12, pp. 1-9, 1982.
- [31] T. Kitamura, S. Imai: "Spectral distortion and quality of synthesized speech in cepstral speech analysis-synthesis system", Electronics and Commun. in Japan, vol. 65-A, Nº 5, pp. 30-38, 1982.
- [32] A. Gray, J. Markel: "Distance measures for speech processing", IEEE Trans. ASSP, vol. 24, Nº 2, pp. 380-391, 1976.
- [33] T. Applebaum, B. Hanson, H. Wakita: "Weighted cepstral distance measures in vector quantization based speech recognizers", Proc. ICASSP, pp. 1155-1158, 1987.
- [34] S. Furui: "Speaker-independent isolated word recognition using dynamic features of speech spectrum", IEEE Trans. ASSP, vol. 34, Nº 1, pp. 52-59, 1986.
- [35] M. Copperi: "Rule-based speech analysis and application to CELP coding", Proc. ICASSP, pp. 143-146, 1988.
- [36] S. Furui: "On the role of spectral transition for speech perception", JASA vol. 80, Nº 4, pp. 1016-1025, 1986.