



**Τμήμα Τηλεπικοινωνιακών Συστημάτων και Δικτύων**

**Τεχνολογικό Εκπαιδευτικό Ίδρυμα Μεσολογγίου**



**ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΜΕΣΟΛΟΓΓΙΟΥ  
ΤΜΗΜΑ ΤΗΛΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ & ΔΙΚΤΥΩΝ**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

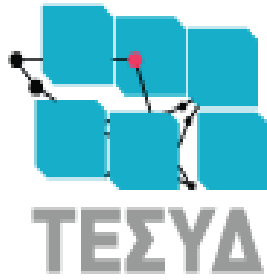
**"Σύστημα Hardware για Υποκειμενική και Αλγοριθμική Αξιολόγηση"**

**Μάντης Ν. Γεώργιος  
Α.Μ: 00197**

Επιβλέπων: Μαριάτος Ευάγγελος

Ναύπακτος 2013





Εγκρίθηκε από την τριμελή εξεταστική επιτροπή:  
Ναύπακτος 25 Απριλίου 2013

#### ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Μαριάτος Ευάγγελος
2. Βώρος Νικόλαος
3. Μαράτου Βίκυ

Απαγορεύεται η αντιγραφή ή αναπαραγωγή ολόκληρου ή τμήματος του κειμένου χωρίς την  
έγγραφη άδεια του δημιουργού.  
All rights reserved. Copyright 2013



*...στη μνήμη της αγαπημένης μου γιαγιάς Δήμητρας*



## Περίληψη

Σκοπός αυτής της εργασίας είναι η ανάλυση των μεθόδων αξιολόγησης όσον αφορά την ποιότητα ομιλίας που λαμβάνει ο τελικός χρήστης της τηλεφωνικής σύνδεσης όπως και οποιοδήποτε άλλου μέσου μεταφοράς της. Για την κωδικοποίηση της φωνής είναι αναγκαίοι οι κωδικοποιητές φωνής, η αξιολόγηση των οποίων γίνεται με τεχνικά εργαλεία ή ακόμη και από ανθρώπους οι οποίοι κρίνουν την ποιότητα του ήχου που τελικά λαμβάνει ο τελικός χρήστης.

Στα κεφάλαια που θα ακολουθήσουν θα γίνει ανάλυση των κωδικοποιητών αυτής της κατηγορίας ώστε να γίνει κατανοητό τι ακριβώς εξυπηρετούν και ποιος είναι ο σκοπός τους.

Συγκεκριμένα στο πρώτο κεφάλαιο γίνεται ανάλυση των κωδικοποιητών φωνής και αναφέρονται τα βασικότερα μοντέλα αυτών. Για να αξιολογηθεί η απόδοση ενός κωδικοποιητή φωνής όμως είναι απαραίτητο να υπάρχει ένας κοινά αποδεκτός δείκτης της ποιότητας ομιλίας, που τελικά αντιλαμβάνεται ο τελικός χρήστης, ο οποίος αντιστοιχεί σε μια κλίμακα αξιολόγησης, η οποία αναλύεται επαρκώς στο δεύτερο κεφάλαιο. Αυτή η κοινώς αποδεκτή κλίμακα αξιολόγησης όμως πέρα από τους κωδικοποιητές είναι και το βασικό εργαλείο αναφοράς των μεθόδων αξιολόγησης που χωρίζονται σε δυο βασικές κατηγορίες.

Κεντρικό σημείο αναφοράς της εργασίας αυτής λοιπόν, είναι οι τεχνικές αξιολόγησης της απόδοσης ενός κωδικοποιητή φωνής αυτής της κλίμακας και γενικότερα της ποιότητας που λαμβάνεται από τον χρήστη. Οι τεχνικές αυτές αναλύονται επακριβώς στα επόμενα δυο κεφάλαια αναφέροντας τον τρόπο λειτουργίας τους και όλες τις τεχνικές προδιαγραφές που αυτές απαιτούν για να εφαρμοστούν όπως ορίζουν τα διεθνή πρότυπα. Τέλος γίνεται συγκριτική μελέτη αυτών των μεθόδων αξιολόγησης βάση των τεχνικών προδιαγραφών τους και του σκοπού για τον οποίο δημιουργήθηκαν, χρησιμοποιώντας τεχνικά συγκριτικά πειράματα διαφόρων ερευνών για την επαλήθευσή της.





## **Abstract**

The object of this work is the analysis of evaluation methods in terms of voice quality received by the end user of the telephone connection as well as any other means of transport. For voice coding is necessary voice codecs, the assessment of which is made with technical tools or even by people who judge the quality of sound that eventually takes the end user.

In the chapters that follow will analyze the coders of this class to understand what and who are serving their purpose. Specifically, in the first chapter analyzes the vocoders and reported the basic models. To evaluate the performance of a vocoder however needs to be a commonly accepted indicator of voice quality, which ultimately understands the end user, that is an assessment scale, which sufficiently analyzed in the second chapter. This commonly accepted rating scale but over codecs is the basic reference tool of evaluation methods are divided into two main categories.

Central points of this work, therefore, are the technical evaluation of a vocoder performance of this scale and quality generally taken by the user. These techniques are analyzed sufficiently in the next two chapters describing how they work and all the technical specifications they require to apply as set by international standards. Finally, is getting a comparative study of these evaluation methods based on the technical specifications and the purpose for which they were created, using technical comparative experiments of various investigations for its verification.



## **Ευχαριστίες**

Ολοκληρώνοντας την παρούσα εργασία, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή κ. Ευάγγελο Μαριάτο, για την ανάθεση της εργασίας και την πολύτιμη βοήθεια και καθοδήγηση του κατά την εκπόνησή της, η οποία ήταν καθοριστική για την πορεία της Πτυχιακής Εργασίας μου. Θα ήθελα επίσης να ευχαριστήσω τους κ. Μαριάτο Ευάγγελο, κ. Βώρο Νικόλαο και κα Μαράτου Βίκυ για την τιμή που μου έκαναν ως μέλη της τριμελούς επιτροπής εξέτασης.

Επίσης θα ήθελα να ευχαριστήσω τους γονείς μου, την αδελφή μου και τους αγαπημένους μου φίλους για την αμέριστη βοήθεια και υποστήριξη που ο καθένας τους μου προσέφερε για την επιτυχή ολοκλήρωση των πτυχιακών μου σπουδών.

Μάντης Ν. Γεώργιος



## Περιεχόμενα

Περίληψη	Σελ. 7
Abstract	Σελ. 9
Ευχαριστίες	Σελ. 11
Περιεχόμενα	Σελ. 13
Πίνακες και σχήματα	Σελ. 15
Συντομογραφίες	Σελ. 17
Πρόλογος	Σελ. 16
<b>1 ΚΩΔΙΚΟΠΟΙΗΤΕΣ ΦΩΝΗΣ</b>	<b>Σελ. 19</b>
1.1 Εισαγωγικές έννοιες και τεχνολογίες υποδομής στους κωδικοποιητές (codecs)	Σελ. 19
1.1.1 Βασικές Αρχές Κωδικοποιητών Φωνής	Σελ. 20
1.1.2 Μοντέλα κωδικοποίησης φωνή	Σελ. 21
1.1.3 Παράμετροι της απόδοσης	Σελ. 22
1.2 Κατηγορίες κωδικοποιητών φωνής	Σελ. 24
1.2.1 Πίνακας	Σελ. 26
1.3 Ανάλυση Κωδικοποιητών Φωνή	Σελ. 28
1.3.1 AMR - AMR-WB (G.722.2) – EVRC	Σελ. 28
1.3.2 G.711, G.721, G.722, G.723, G.726, G.727, G.728 & G.729	Σελ. 29
1.4 Αξιολόγηση της απόδοσης των κωδικοποιητών φωνής	Σελ. 34
<b>2 ΔΕΙΚΤΕΣ ΠΟΙΟΤΗΤΑΣ ΦΩΝΗΣ</b>	<b>Σελ. 36</b>
2.1 Εισαγωγικές έννοιες στους δείκτες ποιότητας ομιλίας	Σελ. 36
2.1.1 Ορισμός ποιότητας εμπειρίας & ποιότητας υπηρεσίας (QoE & QoS)	Σελ. 36
2.1.2 Μέθοδοι αξιολόγησης φωνής	Σελ. 39
<b>3 ΥΠΟΚΕΙΜΕΝΙΚΗ ΜΕΘΟΔΟΣ ΑΞΙΟΛΟΓΗΣΗΣ</b>	<b>Σελ. 42</b>
3.1 MOS (Mean Opinion Score)	Σελ. 42
3.1.1 Δοκιμές Αξιολόγησης Συνομιλίας	Σελ. 44
3.1.2 Δοκιμές Αξιολόγησης Ακρόαση	Σελ. 47
3.1.3 Δοκιμές Δημοσκοπικής Έρευνας	Σελ. 52
<b>4 ΑΝΤΙΚΕΙΜΕΝΙΚΗ-ΑΛΓΟΡΙΘΜΙΚΗ ΜΕΘΟΔΟΣ ΑΞΙΟΛΟΓΗΣΗΣ</b>	<b>Σελ. 55</b>
4.1 Παρεμβατικές αντικειμενικές μέθοδοι	Σελ. 55
4.1.1 Επίδοση του αλγόριθμου PSQM	Σελ. 56
4.1.2 Επίδοση του αλγόριθμου PEAQ	Σελ. 58
4.1.3 Επίδοση του αλγόριθμου PESQ	Σελ. 60
4.1.4 Επίδοση του αλγόριθμου POLQA	Σελ. 66
4.2 Μη Παρεμβατικές αντικειμενικές μέθοδοι	Σελ. 69
4.2.1 Επίδοση του αλγόριθμου 3SQM	Σελ. 70
4.2.2 Επίδοση του αλγόριθμου E-model	Σελ. 72
4.2.3 Επίδοση του αλγόριθμου NiQA	Σελ. 76
<b>5 ΤΕΛΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ – ΣΥΓΚΡΙΣΗ</b>	<b>Σελ. 80</b>
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b>	<b>Σελ. 92</b>



## Πίνακες και σχήματα

Πίνακας 1: Σημαντικότερα πρότυπα ITU	Σελ. 25
Πίνακας 2: General Information & Technical Details of ITU-T Audio Compression Format	Σελ. 26
Πίνακας 3: Χαρακτηριστικά προτύπων κωδικοποιητών μέσω της MOS	Σελ. 34
Πίνακας 4: Κλίμακα γνωμοβαθμολογίας	Σελ. 44
Πίνακας 5: Παραδείγματα από κείμενο ομιλίας	Σελ. 50
Πίνακας 6: Κατηγορίες της ποιότητας μετάδοσης ομιλίας	Σελ. 75
Εικόνα 1: Πρότυπα κωδικοποιητών G.711	Σελ. 30
Εικόνα 2	Σελ. 38
Εικόνα 3	Σελ. 40
Εικόνα 4	Σελ. 40
Εικόνα 5	Σελ. 45
Εικόνα 6	Σελ. 51
Εικόνα 7	Σελ. 58
Εικόνα 8	Σελ. 59
Εικόνα 9	Σελ. 62
Εικόνα 10	Σελ. 63
Εικόνα 11	Σελ. 64
Εικόνα 12	Σελ. 71
Εικόνα 13	Σελ. 88
Εικόνα 14	Σελ. 89
Σχήμα 1: QoS and QoE Parameters – Mapping Model	Σελ. 43
Σχήμα 2	Σελ. 58

Σχήμα 3	Σελ. 61
Σχήμα 4	Σελ. 63
Σχήμα 5: PESQ	Σελ. 65
Σχήμα 6: Evolution of ITU-T Recommendations for Voice Quality Testing	Σελ. 68
Σχήμα 7: E-model	Σελ. 73
Σχήμα 8	Σελ. 77
Γράφημα 1	Σελ. 87
Γράφημα 2	Σελ. 87
Γράφημα 3	Σελ. 88
Γράφημα 4	Σελ. 90



## Συντομογραφίες

ACELP = Algebraic Code-Excited Linear Prediction

ADPCM = Adaptive Differential Pulse Code Modulation

CDMA: Code Division Multiple Access

DCT = Discrete Cosine Transformation

DTMF = Dual-Tone Multi Frequency

ETSI = European Telecommunications Standards Institute

GSM = Global System for Mobile

IRS = Intermediate Reference System

IP = Internet Protocol

ITU = International Telecommunication Union

LD-CELP = Low-Delay Code Excited Linear Prediction

LPC = Linear Prediction Coding

MOS = Mean Opinion Score

MOVs = Multi-Code Output Variables

NIQA = Non-Intrusive voice Quality Analyzer

ODG = Objective Difference Grade

PEAQ = Perceptual Evaluation of Audio Quality

PESQ = Perceptual Evaluation of Speech Quality

PCM = Pulse Code Modulation

POLQA = Perceptual Objective Listening Quality Assessment

POTS = Plain Old Telephone Service

PSQM = Perceptual Speech Quality Measure

QoE = Quality of Experience

QoP = Quality of Perception

QoS = Quality of Service

QoSE = Quality of Service Experience

SDG = Subjective Difference Grade

SNR = Signal to Noise Ratio

UMTS = Universal Mobile Telecommunications System

VAD = Voice Activity Detection

VoIP = Voice over Ip

3 GPP = 3<sup>rd</sup> Generation Partnership Project

3 SQM = Single Speech Quality Measure

## Πρόλογος

Οι κωδικοποιητές ομιλίας είναι μια μέθοδος συμπίεσης/αποσυμπίεσης ενός αρχείου που περιέχει δεδομένα ομιλίας ή μορφή ροής ομιλίας. Η επιλογή της χρήσης ενός κωδικοποιητή γίνεται ανάλογα με τις συνθήκες στις οποίες αυτός θα χρησιμοποιηθεί είτε αυτή είναι μετάδοση είτε αυτή είναι αποθήκευση. Η κωδικοποίηση όμως αυτή πολλές φορές στην ουσία παραμορφώνει τον ήχο και σαν αποτέλεσμα ακούμε κάτι διαφορετικό από αυτό που αρχικά ειπώθηκε. Πέρα από την κωδικοποίηση όμως η φωνή μπορεί να παραμορφωθεί και από τα ίδια τα τηλεφωνικά δίκτυα λόγω της πολυπλοκότητας τους, του εύρους λειτουργίας τους και άλλων αστάθμητων παραγόντων. Στο σημείο αυτό και για να γίνει επιτευκτό ένα άρτιο αποτέλεσμα μεταφοράς του ήχου, όσο το δυνατόν με τις λιγότερες απώλειες ή παραμορφώσεις, είναι απαραίτητες οι μέθοδοι αξιολόγησης όπου τα αποτελέσματά τους είναι απολύτως απαραίτητα για την συνεχή εξέλιξη και αναβάθμιση των δικτύων. Οι αξιολογήσεις αυτές γίνονται είτε από ανθρώπους είτε από αλγόριθμους. Και οι δυο μέθοδοι έχουν κοινή κλίμακα αξιολόγησης την λεγόμενη MOS από την οποία μπορούμε να εξαγάγουμε εύκολα συμπεράσματα ως προς την παρεχόμενη ποιότητα από τα δίκτυα τηλεφωνίας. Στα κεφάλαια που ακλουθούν θα αναφερθούν λεπτομερώς όλοι οι βασικοί κωδικοποιητές που χρησιμοποιούνται σήμερα όπως και θα αναλυθεί η κοινή μέθοδος βαθμολόγησης και τα κριτήρια αυτής. Επίσης θα αναλυθούν όλες οι βασικές μέθοδοι υποκειμενικής και αλγοριθμικής αξιολόγησης με τις τεχνικές προδιαγραφές που ορίζει η ITU για αυτές τις μεθόδους. Τέλος θα γίνει συγκριτική αξιολόγηση όλων των μεθόδων με αναφορά σε πειράματα μερικών από αυτών.



# ΚΕΦΑΛΑΙΟ 1

## ΚΩΔΙΚΟΠΟΙΗΤΕΣ ΦΩΝΗΣ

### Εισαγωγή

#### 1.1 Εισαγωγικές έννοιες και τεχνολογίες υποδομής στους κωδικοποιητές (codecs)

Στη εποχή της επικοινωνίας με την ευρύτατη χρήση του τηλεφωνικού δικτύου, η μετάδοση φωνής με αποδοτικό τρόπο έχει τεράστια οικονομική σημασία, ιδιαίτερα σε κανάλια μετάδοσης με περιορισμένο εύρος ζώνης, όπως για παράδειγμα τα δίκτυα τηλεφωνίας. Από την άλλη πλευρά, η αποδοτική κωδικοποίηση της μουσικής με υψηλή ποιότητα είναι κρίσιμη για τη μουσική βιομηχανία, ενώ η κωδικοποίηση φωνής, μουσικής και άλλων ήχων είναι πολύ σημαντική για την κινηματογραφική βιομηχανία. Το αποτέλεσμα είναι ότι έχουν αναπτυχθεί πολλές μέθοδοι κωδικοποίησης ήχου είτε ειδικά για την ανθρώπινη φωνή, είτε γενικότερα για μουσικά ηχητικά σήματα. Σε όλες αυτές τις μεθόδους, αποσκοπούμε να δώσουμε στον ακροατή την καλύτερη δυνατή αναπαράσταση του ήχου που παράγεται από τον αποστολέα, με όσο το δυνατόν χαμηλότερη κατανάλωση εύρους ζώνης. Ανεξάρτητα από τη φύση του ήχου, σε όλες τις τεχνικές αρχικά γίνεται δειγματοληψία του σήματος μέσω της μεθόδου PCM (Pulse Code Modulation) με τον κατάλληλο ρυθμό δειγματοληψίας και πλήθος bit ανά δείγμα, και στη συνέχεια τα δείγματα αναλύονται για να μπορέσουν να κωδικοποιηθούν με μικρότερο αριθμό bit. Η βασική λειτουργία των κωδικοποιητών λοιπόν είναι η κωδικοποίηση των δειγμάτων φωνής μέσω της μεθόδου PCM σε αριθμό από bits έτσι ώστε η αναπαραγωγή της φωνής να δικαιολογεί την ύπαρξη πιθανόν σφαλμάτων και καθυστέρησης της εκπομπής του ήχου.

Η επιλογή της χρήσης ενός κωδικοποιητή γίνεται ανάλογα με τις συνθήκες στις οποίες αυτός μπορεί να χρησιμοποιηθεί είτε αυτή είναι μετάδοση είτε είναι αποθήκευση. Για την περίπτωση της μετάδοσης μας ενδιαφέρει ο κωδικοποιητής να εισάγει όσο το δυνατόν λιγότερη καθυστέρηση ειδικά όταν σε αυτή υπάρχουν διάφορες επιπρόσθετες καθυστερήσεις, ενώ για την αποθήκευση η παράμετρος της καθυστέρησης δεν είναι μεγάλης σημασίας. Κάθε κωδικοποιητής έχει επίσης διαφορετικές δυνατότητες συμπίεσης.

Στις επόμενες ενότητες θα δούμε πληρέστερα τις βασικές αρχές των κωδικοποιητών και θα αναλύσουμε τους βασικότερους.

### **1.1.1 Βασικές αρχές κωδικοποιητών φωνής**

Για την μεταφορά των αναλογικών σημάτων φωνής σε ένα ψηφιακό δίκτυο απαραίτητη η μετατροπή τους σε ψηφιακή μορφή. Αυτό γίνεται με την βοήθεια των αλγορίθμων κωδικοποίησης της φωνής, δηλαδή τους αποκωδικοποιητές. Ένας κωδικοποιητής παίρνει ως είσοδο το αναλογικό σήμα της φωνής και το κωδικοποιεί δημιουργώντας πλαίσια που έχουν το ίδιο μέγεθος. Χρησιμοποιεί επίσης τεχνικές συμπίεσης και αποσυμπίεσης της φωνής, με σκοπό να μειωθεί το απαιτούμενο εύρος ζώνης για την μεταφορά της. Όσο μεγαλύτερη συμπίεση πραγματοποιείται τόσο λιγότερο εύρος ζώνης χρησιμοποιείται με αποτέλεσμα να μειώνεται η ποιότητα της συνομιλίας. Υπάρχουν διάφοροι κωδικοποιητές που χρησιμοποιούνται οι οποίοι διαφέρουν στον αλγόριθμο κωδικοποίησης, στο απαιτούμενο εύρος ζώνης, στην ποιότητα της φωνής και στην πολυπλοκότητα. Μερικοί κωδικοποιητές υποστηρίζουν επίσης την τεχνική ανίχνευσης δραστηριότητας φωνής (Voice Activity Detection).

### 1.1.2 Μοντέλα κωδικοποίησης φωνής

Τα βασικά μοντέλα παραγωγής ομιλίας στα οποία βασίζεται η κωδικοποίηση της φωνής είναι το γνωστικό, το γλωσσικό, το μοντέλο φωνητικού σωλήνα και τέλος το ακουστικό μοντέλο ή μοντέλο κρίσιμης ζώνης.

- i) **Γνωστικό Μοντέλο:** Βασίζεται στις διεργασίες που λαμβάνουν χώρα στον εγκέφαλο για την παραγωγή της ομιλίας. Το μοντέλο αυτό βρίσκεται ακόμη στο στάδιο της έρευνας και δεν έχει εφαρμοσθεί σε ένα σύστημα κωδικοποιητή ομιλίας.
- ii) **Γλωσσικό Μοντέλο:** Ο προφορικός λόγος παρουσιάζει ολοκληρωμένη συντακτική δομή η οποία σχετίζεται με ένα μοτίβο συλλαβικής έντασης. Σε αυτό βασίζεται το γλωσσικό μοντέλο. Στον προφορικό λόγο δεν υπάρχει μόνο η πληροφορία για το νοηματικό περιεχόμενο αλλά και η στάση του ομιλητή σε αυτό. Οι κωδικοποιητές που κάνουν χρήση αυτού του μοντέλου παράγουν ομιλία χαμηλής ποιότητας αλλά μας δίνουν την δυνατότητα να απομονώσουμε στοιχεία της ομιλίας (για παράδειγμα το νόημα της πρότασης) και έτσι να μπορέσουμε να εστιάσουμε αποκλειστικά στα χαρακτηριστικά της κυματομορφής της ομιλίας που μας ενδιαφέρουν.
- iii) **Μοντέλο Φωνητικού Σωλήνα:** Το μοντέλο αυτό βασίζεται στον τρόπο με τον οποίο λειτουργεί η φωνητική περιοχή του ανθρώπου και στους βασικούς ήχους που αυτή μπορεί να παράγει, έμφωνα ή άφωνα. Οι κωδικοποιητές που χρησιμοποιούν αυτό το μοντέλο προσπαθούν να μοντελοποιήσουν την διεργασία παραγωγής της φωνής με ένα δυναμικό σύστημα και επιπλέον προσπαθούν να ποσοτικοποιήσουν συγκεκριμένους περιορισμούς γι' αυτό το σύστημα. Βασικές λειτουργίες αυτών των κωδικοποιητών είναι να αναλύουν το σήμα της ομιλίας στον πομπό, να μεταδίδουν τις παραμέτρους που προκύπτουν από την ανάλυση και έπειτα χρησιμοποιώντας τις παραμέτρους αυτές να επανασυνθέτουν την ομιλία στο

δέκτη. Οι κωδικοποιητές αυτοί αποδίδουν μικρό αριθμό bits και η ποιότητα δεν είναι πάντοτε καλή.

iv) Ακουστικό Μοντέλο ή Μοντέλο Κρίσιμης Ζώνης: Αν οποιαδήποτε πληροφορία στο αυθεντικό σήμα ομιλίας φιλτραριστεί και απορριφθεί από το ανθρώπινο ακουστικό σύστημα, τότε η πληροφορία μπορεί να αφεθεί εκτός της κωδικοποιημένης αναπαράστασης του σήματος, με αποτέλεσμα τα bits που είναι απαραίτητα για να αναπαρασταθεί το σήμα να μειωθούν. Ένα μοντέλο που έχει χρησιμοποιηθεί από τους κωδικοποιητές ομιλίας με επιτυχία είναι το ακουστικό ή κρίσιμης ζώνης για την ακουστική αντίληψη. Το μοντέλο αυτό προσπαθεί να συγκρατήσει έναν αριθμό συσχετιζόμενων στοιχείων της ακουστικής αντίληψης. Το πρώτο στοιχείο είναι η ανάλυση ακουστικής συχνότητας. Το εύρος μιας κρίσιμης ζώνης σε μια συγκεκριμένη συχνότητα είναι το μέτρο του πόσο απομακρυσμένοι πρέπει να είναι δύο τόνοι στη συχνότητα ώστε να είναι διακριτοί μεταξύ τους. Το δεύτερο στοιχείο είναι η ακουστική επικάλυψη θορύβου. Αυτό το στοιχείο διατυπώνεται με το εξής παράδειγμα, ένα σήμα ομιλίας σε μια συγκεκριμένη κρίσιμη ζώνη θα επικαλύψει τον θόρυβο που βρίσκεται στην ίδια ζώνη. Έτσι σήματα θορύβου που βρίσκονται κοντά στο σήμα ομιλίας στο πεδίο της συχνότητας καλύπτονται.

### **1.1.3 Παράμετροι της απόδοσης**

Οι κωδικοποιητές προσπαθούν να ελαχιστοποιήσουν τον ρυθμό bit για την μετάδοση ή την αποθήκευση του σήματος, διατηρώντας παράλληλα τα απαιτούμενα επίπεδα ποιότητας της ομιλίας, της καθυστέρησης της επικοινωνίας και της πολυπλοκότητας της εφαρμογής. Οι παράμετροι αυτής απόδοσης αυτής είναι:



- i) Η Ποιότητα της Ομιλίας: Αξιολογείται συνήθως σε μια πενταβάθμια κλίματα, που είναι γνωστή ως MOS (Mean Opinion Score) για τον έλεγχο της ποιότητας ομιλίας. Τα πέντε σημεία της ποιότητας είναι τα εξής: κακή, χαμηλή, ικανοποιητική, καλή και εξαιρετική. Η βαθμολογία με αποτελέσματα άνω του 3,5 ή υψηλότερη σημαίνει γενικά υψηλά επίπεδα κατανόησης, την αναγνώριση του ομιλητή και φυσικότητα της φωνής.
- ii) Ο ρυθμός των bit: Η αποτελεσματικότητα της κωδικοποίησης εκφράζεται σε bits ανά δευτερόλεπτο (bps).
- iii) Η Καθυστέρηση της Επικοινωνίας: Οι κωδικοποιητές ομιλίας συχνά επεξεργάζονται τα δείγματα σε block, και έτσι εμφανίζεται μια χρονική καθυστέρηση μεταξύ του αυθεντικού και του κωδικοποιημένου σήματος ομιλίας. Ανάλογα με την εφαρμογή υπάρχουν αυστηροί περιορισμοί στην επιτρεπόμενη συνολική καθυστέρηση.
- iv) Η Πολυπλοκότητα: Είναι η προσπάθεια επεξεργασίας που απαιτείται για την υλοποίηση του αλγορίθμου.
- v) Η Αντοχή στους Θορύβους τους Περιβάλλοντος: Είναι η παράμετρος που μας δείχνει πόσο καλά μπορεί να λειτουργήσει ο αλγόριθμος κωδικοποίησης σε ένα περιβάλλον με θόρυβο.
- vi) Η Αντοχή στα Σφάλματα: Αυτή η παράμετρος είναι η πιο σημαντική γιατί πολλοί κωδικοποιητές χρησιμοποιούνται σε ασύρματες επικοινωνίες. Σε τέτοια συστήματα, οι κωδικοποιητές πρέπει να παρέχουν ικανοποιητική ποιότητα ομιλίας με ρυθμό σφαλμάτων έως και 10%.

## 1.2 Κατηγορίες κωδικοποιητών φωνής

Οι κωδικοποιητές ανάλογα με τα χαρακτηριστικά του σήματος της ομιλίας (έμφωνη ή άφωνη ομιλία) και σε συνδυασμό με την χρήση κάποιου μοντέλου κωδικοποίησης φωνής κατηγοριοποιούνται σε τρεις διαφορετικούς τύπους: τους κωδικοποιητές κυματομορφής (waveform coding), τους κωδικοποιητές φωνής (vocoding) και τους υβριδικούς κωδικοποιητές (hybrid coding), συνδυάζονται οι δύο παραπάνω κατηγορίες.

Οι Κωδικοποιητές Κυματομορφής (Waveform Coding) αναπαράγουν την αναλογική κυματομορφή με τη μεγαλύτερη δυνατή ακρίβεια συμπεριλαμβανομένου και του θορύβου. Με δεδομένο ότι ενεργούν απευθείας στα αναλογικά σήματα εισόδου παράγονται δείγματα υψηλής ποιότητας. Παρ' όλα αυτά παράγουν δείγματα με υψηλό ρυθμό bits/s. Για παράδειγμα ο ITU-T G.711 χρησιμοποιεί ρυθμό 64 kbits/s.

Οι Κωδικοποιητές Φωνής (Vocoding) έχουν σχεδιαστεί για εφαρμογές χαμηλού bit rate και ο βασικός τους στόχος είναι να διατηρήσουν την κατανοησιμότητα της ομιλίας. Οι περισσότεροι βασίζονται στην κωδικοποίηση γραμμικής εκτίμησης LPC (Linear Prediction Coding) μέσω της οποίας κάθε πλαίσιο ομιλίας μοντελοποιείται ως η έξοδος ενός γραμμικού συστήματος που αναπαριστά την φωνητική περιοχή, σε ένα σήμα διέγερσης. Επιτυγχάνουν χαμηλότερο bit rate σε σχέση με τους κωδικοποιητές κυματομορφής με ρυθμούς κάτω από 2 kb/s.

Οι Υβριδικοί Κωδικοποιητές (Hybrid Coding) συνδυάζουν τα χαρακτηριστικά των δύο παραπάνω κατηγοριών και παρουσιάζουν ορισμένα πλεονεκτήματα. Έχουν την ικανότητα της αποτελεσματικής κωδικοποίησης των κωδικοποιητών φωνής και την ποιότητα του σήματος των κωδικοποιητών κυματομορφής και λειτουργούν σε χαμηλούς ρυθμούς bit rate (4-16 kbits).

Τα σημαντικότερα πρότυπα της ITU (International Telecommunications Union) για την κωδικοποίηση φωνής που αντικατοπτρίζουν τις απαιτήσεις για αποτελεσματική μετάδοση από διάφορα τηλεπικοινωνιακά κανάλια είναι τα ακόλουθα:

ITU-T Designation	Εύρος ζώνης (Hz)	Συχνότητα Δειγματοληψίας (Hz)	Bit Rate (kbits/s)
<b>G. 711</b>	200 – 3200	8	64
<b>G. 722</b>	50 – 7000	16	64
<b>G. 721</b>	200 – 3200	8	32
<b>G. 728</b>	200 – 3200	8	16
<b>G. 723</b>	200 – 3200	8	5,3 & 6,3

**Πίνακας 1: Σημαντικότερα Πρότυπα ITU**

Αυτά τα μοντέλα κωδικοποιητών ομιλίας διαφοροποιούνται τεχνικά μεταξύ τους με βάση διάφορους παράγοντες που περιλαμβάνουν την τεχνολογία συμπίεσης ή τον αλγόριθμο, το εύρος ζώνης, τα ποσοστά των δεδομένων κ.α.

Η καταλληλότητα του κωδικοποιητή που πρέπει να χρησιμοποιηθεί εξαρτάται και από την εφαρμογή. Ωστόσο, τα πλεονεκτήματα και τα μειονεκτήματα κατανόησης ορισμένων από αυτούς τους κωδικοποιητές μα δίνουν καλύτερη πληροφόρηση και διορατικότητα σε βάθος.

Στις παραγράφους που ακολουθούν θα παρουσιαστούν επίσης τα εξής πρότυπα: G.711.0, G.711.1, G722.2 (AMR-WB), G.722.1, G.723.1, G.726, G.727, G.728, G.729 & AMR.

## 1.2.1 Πίνακας

Στον πίνακα που ακολουθεί παρουσιάζονται συγκεντρωμένα πληροφορίες για τα πρότυπα των κωδικοποιητών που θα δούμε αναλυτικά στην επόμενη ενότητα.

### General Information & Technical Details of ITU-T Audio Compression Formats

Audio Compression Format	Creator	Description	First Public Release Date	Patented	Algorithm	Sample Rate	Bit Rate	Bits for sample	Latency
AMR	3 GPP		22/06/1999	Non-free Yes					
G.711	ITU-T	Pulse Code Modulation (PCM)	ITU-T standard from 1988	No	companding A-law or $\mu$ -law, PCM, Lossy	8 kHz	64 kbit/s	8 bit	125 ms
G.711.1	ITU-T				MDCT, A-law, $\mu$ -law, Lossy	8 or 16 kHz	64, 80, 96 kbit/s	16 bit	11.875 ms
G.721	ITU-T	Adaptive Differential Pulse Code Modulation (ADPCM)			ADPCM, Lossy	8 kHz	32 kbit/s	13 bit	
G.722	ITU-T	7 kHz audio-coding within 64 kbit/s	1988-11	No	sub-band ADPCM, Lossy	16 kHz	64 kbit/s	14 bit	4 ms
G.722.1	ITU-T	Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss	1999-09	Non-free Yes	Modulated Lapped Transform, (based on Siren Codec), Lossy	16 kHz	24, 32 kbit/s	16 bit	40 ms
G.722.2 (AMR-WB)	3 GPP		10/04/2001	Non-free Yes	multi-rate wideband ACELP, Lossy	16 kHz	6.60–23.85 kbit/s	14 bit	25 ms
G.723	ITU-T	Extensions of Recommendation G.721 adaptive differential pulse code modulation to 24 and 40			ADPCM, Lossy	8 kHz	24 or 40 kbit/s	13 bit	

		kbit/s for digital circuit multiplication equipment application							
G.723.1	ITU-T	Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s	1996-03	Non-free Yes	MP-MLQ, ACELP, Lossy	8 kHz	5.3, 6.3 kbit/s	13 bit	37.5 ms
G.726	ITU-T	40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)	1990-12	No	ADPCM, Lossy	8 kHz	16, 24, 32, 40 kbit/s	13 bit	125 ms
G.728	ITU-T	Coding of speech at 16 kbit/s using low-delay code excited linear prediction	1992-09	Non-free Yes	low-delay CELP, Lossy	8 kHz	16 kbit/s	13 bit	0.625 ms
G.729	ITU-T	Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)	1996-03	Non-free Yes	CS-ACELP, Lossy	8 kHz	8 kbit/s	13 bit	15 ms

**Πίνακας 2**

### 1.3 Ανάλυση κωδικοποιητών φωνής

Σε αυτή την ενότητα θα δούμε αναλυτικά τα βασικότερα χαρακτηριστικά των δημοφιλέστερων κωδικοποιητών

### 1.3.1 AMR - AMR-WB (G.722.2) – EVRC

- **Adaptive Multi-Rate (AMR)**

Είναι ένα ηχητικό σύστημα συμπίεσης που βελτιστοποιεί τα δεδομένα των κωδικοποιητών ομιλίας. Το AMR καθιερώθηκε ως πρότυπο κωδικοποιητή ομιλίας του 3 GPP τον Οκτώβριο του 1999 και σήμερα χρησιμοποιείται ευρέως σε δίκτυα GSM και UMTS. Πολλά σύγχρονα κινητά τηλέφωνα μπορούν να αποθηκεύσουν σύντομες ηχογραφήσεις σε μορφή AMR. Είναι κατάλληλο για εφαρμογές με πιο προχωρημένες απαιτήσεις σχετικά με τη μορφή αποθήκευσης, όπως η τυχαία πρόσβαση ή ο συγχρονισμός με το βίντεο. Σαν πλεονέκτημα θεωρείται η ανώτερη ποιότητα του ήχου λόγω ενός ευρύτερου εύρους ζώνης ομιλίας. Το μειονέκτημα του είναι η καθυστέρηση που εισάγει στη διαδρομή της φωνής. Κωδικοποιεί σε στενή ζώνη από 4.75 έως 12.2 kbps.

Άλλες εκδόσεις του AMR είναι:

- Adaptive Multi-Rate Wide Band (AMR-WB): είναι ένα πρότυπο κωδικοποίησης ομιλίας που αναπτύχθηκε μετά το AMR χρησιμοποιώντας την ίδια τεχνολογία όπως το ACELP, είναι επίσης γνωστός ως G.722.2. Προσφέρει χαμηλότερο ρυθμό διαμεταγωγής από τον G.722.1 στα 6.6 - 23,85 kbit/s. Μπορεί και προσαρμόζεται γρήγορα στις διάφορες συμπίεσεις και το εύρος ζώνης διατηρείται όταν η συμφόρηση του δικτύου είναι υψηλή.
- Adaptive Multi-Rate Narrow Band (AMR-NB): είναι ένας κωδικοποιητής ομιλίας που χρησιμοποιείται σε χαμηλού ρυθμού εφαρμογές, όπως τα κινητά τηλέφωνα. Είναι επίσης μια εφαρμογή του ACELP.

- **EVRC**

#### Enhanced Variable Rate Codec

Είναι ένας κωδικοποιητής ομιλίας που χρησιμοποιείται σε δίκτυα CDMA. Αναπτύχθηκε το 1995 για να αντικαταστήσει των κωδικοποιητή φωνής QCELP και πρωταρχικός του στόχος ήταν να προσφέρει στους παρόχους κινητής τηλεφωνίας μεγαλύτερη χωρητικότητα των δικτύων τους, ενώ δεν αυξάνεται το ποσό του εύρους ζώνης ή του ασύρματου φάσματος που απαιτείται. Έχει τη δυνατότητα να συμπίεζει κάθε 20 χιλιοστά του δευτερολέπτου 8000 Hz. Ο μέσος ρυθμός διαμεταγωγής ποικίλει με βάση τις συνθήκες του δικτύου αλλά είναι συνήθως 6 kbit/s.

Ο EVRC αντικαταστάθηκε από τον SMV (Selectable Mode Vocoder), ο οποίος διατήρησε την ποιότητα της ομιλίας και ταυτόχρονα βελτίωσε την χωρητικότητα του δικτύου.

### **1.3.2 G.711, G.721, G.722, G.723, G.726, G.727, G.728, G.729**

- **G.711**

Ο G.711 είναι ένας από τους παλιότερους κωδικοποιητές (κυκλοφόρησε το 1972) και χρησιμοποιεί ένα ρυθμό δειγματοληψίας 8000 δειγμάτων ανά δευτερόλεπτο. Ο κωδικοποιητής αυτός έχει δύο εκδόσεις, την μ-Law η οποία χρησιμοποιείται κυρίως στην Β. Αμερική και την Ιαπωνία και την A-Law η οποία χρησιμοποιείται στην Ευρώπη και τον υπόλοιπο κόσμο. Για την κωδικοποίηση των αναλογικών σημάτων χρησιμοποιεί την παλμοκωδική διαμόρφωση (PCM) με συχνότητα δειγματοληψίας 8 KHz. Το κάθε δείγμα αναπαρίσταται με 8 bit με αποτέλεσμα να έχει ένα ρυθμό μετάδοσης 64kbps. Κάθε πακέτο που δημιουργεί αποτελείται από ένα πλαίσιο το οποίο περιέχει δείγματα διάρκειας 20 ms.

Συνεπώς, το ωφέλιμο φορτίο του κάθε πακέτου που δημιουργεί κάθε 20 ms είναι 160 Byte.

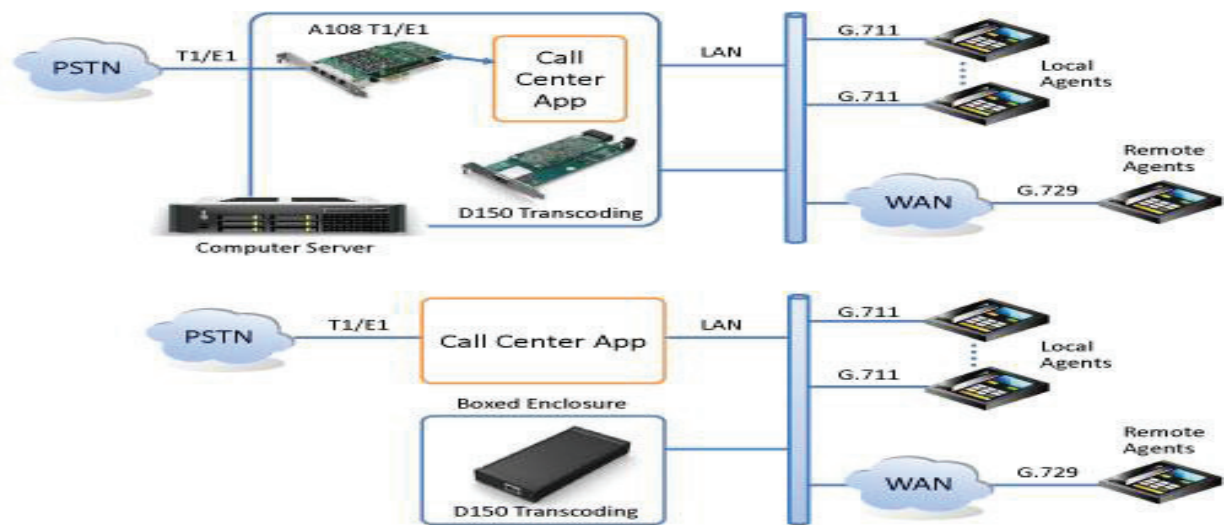
Στα πλεονεκτήματα του μοντέλου G.711 συγκαταλέγονται τα πολύ χαμηλά γενικά έξοδα επεξεργασίας και η σχεδίασή του, που προσφέρει λεπτομερή μετάδοση της ομιλίας.

Έχει όμως και ορισμένα μειονεκτήματα, όπως η κακή απόδοση του δικτύου και η απουσία παρεμβολής χαμένων πακέτων.

Άλλες εκδόσεις του G.711 είναι:

- G.711.0: Είναι επίσης γνωστός ως G.711.0 LLC. Χρησιμοποιεί συμπιεσμένα δεδομένα, χωρίς απώλειες, για την μείωση της χρήσης τους εύρους ζώνης τόσο όσο το 50%. Η παλμοκωδική διαμόρφωση του προτύπου εγκρίθηκε από την ITU-T το Σεπτέμβριο του 2009.

- G.711.1: Είναι μια σύσταση του G.711 που δημοσιεύθηκε τον Μάρτιο του 2008. Ο κωδικοποιητής G.711.1 δημιουργεί ένα ενιαίο σήμα ήχου διαβαθμισμένο σε τρία επίπεδα που αντιστοιχούν σε τρεις διαφορετικούς ρυθμούς διαμεταγωγής: 64, 80 και 96 kbit/s.



Εικόνα 1: Πρότυπα κωδικοποιητών G.711

- **G.721**

Κυκλοφόρησε το 1984 αλλά αντικαταστάθηκε το 1990 από τον G. 726.



- **G.722**

Είναι ένα πρότυπο ITU λειτουργίας ευρείας ζώνης κωδικοποιητή ομιλίας στα 48, 56 & 64 kbit/s. Εγκρίθηκε από την ITU-T τον Νοέμβριο του 1988 και η τεχνολογία του βασίζεται στην υποζώνη ADPCM. Είναι χρήσιμο σε δίκτυα σταθερής τηλεφωνίας μέσω IP εφαρμογών, όπου το απαιτούμενο εύρος ζώνης τυπικά δεν είναι απαγορευτικό. Προσφέρει επίσης μια σημαντική βελτίωση στην ποιότητα της ομιλίας σε παλιούς οριακούς κωδικοποιητές όπως ο G.711. Από την άλλη δεν είναι βέλτιστο για απομακρυσμένη εκπομπή δεδομένων.

Άλλες εκδόσεις είναι:

- G.722.1: είναι ένα ITU-T πρότυπο κωδικοποιητή ήχου που χρησιμοποιείται για ομιλία υψηλής ποιότητας. Ο G.722.1 είναι ένας μετασχηματισμός με βάση συμπίεση που έχει βελτιστοποιηθεί τόσο για ομιλία όσο και για μουσική. Προσφέρει χαμηλό bit rate συμπίεσης στα 24 kbit/s ή στα 32 kbit/s.
- G.722.2: Βλέπε παράγραφο 1.3.1: AMR-WB

- **G.723**

Είναι ένα ITU-T πρότυπο κωδικοποιητή ομιλίας που χρησιμοποιεί επεκτάσεις του G.721 παρέχοντας ποιότητα φωνής που καλύπτει από 300 Hz έως 3400 Hz με την χρήση της μεθόδου ADPCM (Adaptive Differential Pulse Code Modulation) και παρέχει καλή ποιότητα ήχου στα 24 και στα 40 kbps. Αυτός ο κωδικοποιητής κυκλοφόρησε το 1988 αλλά θεωρείται ξεπερασμένος γιατί το 1990 αντικαταστάθηκε από τον G.726.

- **G.723.1**

Είναι ένας κωδικοποιητής διαφορετικός από τον G.723. Υποστηρίζει δύο ρυθμούς μετάδοσης: 5,3 και 6,3 kbps. Η συχνότητα δειγματοληψίας είναι 8 KHz και το κάθε πλαίσιο

περιέχει δείγματα διάρκειας 30 ms. Έχει αρκετά πλεονεκτήματα όπως η δυνατότητα υψηλής συμπίεσης, διατηρώντας όμως ταυτόχρονα υψηλή ποιότητα ήχου και επιτρέπει την ταυτόχρονη κωδικοποίηση και αποκωδικοποίηση σε λογισμικό (σε γρήγορους υπολογιστές). Είναι πολύ αποτελεσματικός στο ηχητικό τμήμα της τηλεδιάσκεψης και της τηλεφωνίας στα δημόσια τηλέφωνα (POTS). Από την άλλη θα λέγαμε ότι απαιτεί πολύ δύναμη από τον επεξεργαστή, δεν είναι κατάλληλο για μουσική ή ηχητικά εφέ και είναι χαμηλής ποιότητας σε σχέση με πολλούς κωδικοποιητές σε παρόμοια ποσοστά δεδομένων.

- **G.726**

Είναι ένας ADPCM κωδικοποιητής ομιλίας για την μετάδοση φωνής σε ποσοστά 16, 24, 32 και 40 kbit/s. Δημιουργήθηκε για να αντικαταστήσει τον G.721 και τον G.723. Η πιο διαδεδομένη έκδοση που χρησιμοποιείται είναι αυτή με ρυθμό 32 kbps, τα οποία είναι το μισό ποσοστό του G.711 και ως εκ τούτου αυξάνει την χωρητικότητα του διαθέσιμου δικτύου κατά 100%. Χρησιμοποιείται επίσης πάρα πολύ σε διεθνή υπεραστικά τηλεφωνήματα στο τηλεφωνικό δίκτυο. Και αυτός ο κωδικοποιητής, όπως και ο G.723.1 δεν είναι κατάλληλος για μουσικά ή ηχητικά εφέ.

- **G.727**

Ο κωδικοποιητής G.727 καθιερώθηκε τον ίδιο χρόνο (1990) με τον G.726 και περιλαμβάνει τον ίδιο ρυθμό διαμεταγωγής αλλά έχει βελτιστοποιηθεί για πακέτα πολλαπλών κυκλωμάτων εξοπλισμού.

- **G.728**

Ο κωδικοποιητής G.728 είναι της τεχνολογίας LD-CELP (Low Delay- Code Excited Linear Prediction) και λειτουργεί με πέντε δείγματα ανά δειγματοληψία. Αξιολογείται σαν «αριθμός ποιότητας», έτσι η ποιότητα της φωνής είναι πολύ καλή σε σύγκριση με τους προηγούμενους κωδικοποιητές ομιλίας. Επιπλέον είναι πολύ μικρής καθυστέρησης, ως εκ τούτου περιλαμβάνει τους δορυφόρους, τα κινητά και τα συστήματα τηλεδιάσκεψης με βίντεο. Αρνητικό ωστόσο θεωρείται ότι είναι διαθέσιμα λίγα bit για προστασία σε περίπτωση σφάλματος.

- **G.729**

Ο κωδικοποιητής ομιλίας G.729 χρησιμοποιεί έναν αλγόριθμο συμπίεσης δεδομένων ήχου και συμπιέζει τα δεδομένα στα ποσοστά που κυμαίνονται μεταξύ 6,4 και 12.4 kbps. Απαιτεί χαμηλή καθυστέρηση για την συμπίεση των δεδομένων ομιλίας, τόσο χαμηλή όσο 10 χιλιοστά του δευτερολέπτου. Αυτό έχει σαν αποτέλεσμα, μουσική ή ήχοι, όπως ήχοι DTMF, ή φαξ να μην μπορούν να μεταφερθούν αξιόπιστα με αυτόν τον κωδικοποιητή. Λόγω και του χαμηλού εύρους ζώνης, περίπου 8 kbps, χρησιμοποιείται ως επί το πλείστον σε VoIP (Voice over IP) εφαρμογές για χαμηλή απαίτηση εύρους ζώνης. Αρνητικό είναι επίσης το γεγονός ότι η ποιότητα της ομιλίας μειώνεται οριακά και απαιτείται άδεια για την χρήση του.

#### **1.4 Αξιολόγηση της απόδοσης των κωδικοποιητών φωνής**

Για να αξιολογήσουμε την απόδοση ενός κωδικοποιητή φωνής είναι απαραίτητο να έχουμε κάποιο δείκτη της κατανοησιμότητας και της ποιότητας ομιλίας που παράγεται. Ο όρος

κατανοησιμότητα αναφέρεται στο αν η παραγόμενη ομιλία είναι εύκολα κατανοητή ενώ η ποιότητα είναι ένας δείκτης που δείχνει πόσο φυσικά ακούγεται η ομιλία.

Οι τεχνικές αξιολόγησης της απόδοσης ενός κωδικοποιητή φωνής όσον αφορά την κατανοησιμότητα και την ποιότητα της ομιλίας που αυτός παράγει χωρίζονται σε δύο κατηγορίες, τις υποκειμενικές και τις αντικειμενικές.

Στον επόμενο πίνακα απεικονίζονται τα χαρακτηριστικά ορισμένων προτύπων κωδικοποίησης φωνής μέσω της υποκειμενικής τεχνικής αξιολόγησης MOS .

Coding Standard	Year	Bit Rate (kbps)	MOS	MIPS	Frame Size (msec)
<b>G.711 μ-Law PCM</b>	1972	64	4.3	0.01	0.125
<b>G.721 ADPCM</b>	1984	32	4.1	2	0.125
<b>G.722 ADPCM</b>	1984	48/56/64	4.1	5	0.125
<b>G.726/G.727 ADPCM</b>	1990	16/24/32/40	4.1	2	0.125
<b>G.728 LD-CELP</b>	1992	16	4	30	0.625
<b>G.729 CS-ACELP</b>	1996	8	4	20	10
<b>G.723.1 MPC-MLQ</b>	1995	6.3/5.3	4/3.7	11	10

**Πίνακας 3: Χαρακτηριστικά Προτύπων Κωδικοποιητών μέσω της MOS**

Στο κεφάλαιο 3 και συγκεκριμένα στην παράγραφο 3.1 θα δούμε αναλυτικά την υποκειμενική τεχνική αξιολόγησης MOS.

## **ΚΕΦΑΛΑΙΟ 2**

### **ΔΕΙΚΤΕΣ ΠΟΙΟΤΗΤΑΣ ΦΩΝΗΣ**

#### **2.1 Εισαγωγικές έννοιες στους δείκτες ποιότητας ομιλίας**

Η έννοιες της ποιότητας υπηρεσίας ή ποιότητας εμπειρίας ή και αντιλαμβανόμενης ποιότητας είναι σχετικά νέοι όροι που δημιουργήθηκαν από την ανάγκη συνεχούς βελτίωσης των προϊόντων που σχετίζονται με τις τηλεπικοινωνίες. Ο λόγος δημιουργίας τους είναι η ύπαρξη ενός σημείου αναφοράς μέσω ενός δείκτη που αντιστοιχεί σε ένα βαθμολογικού πίνακα και ο οποίος θα ορίζει κάθε φορά το κατά πόσο το υπό δοκιμή προϊόν ή υπηρεσία φτάνει στα επιθυμητά επίπεδα λειτουργίας.

##### **2.1.1 Ορισμός ποιότητας εμπειρίας & ποιότητας υπηρεσίας (QoE & QoS)**

Το Ευρωπαϊκό Ινστιτούτο Τηλεπικοινωνιών Προτύπων (ETSI, European Telecommunications Standards Institute) ορίζει σαν ποιότητα υπηρεσίας "Το μέτρο της ικανοποίησης του χρήστη από τη χρήση μιας υπηρεσίας ή ενός προϊόντος τεχνολογίας πληροφορικής και επικοινωνιών βασισμένο σε αντικειμενικά και υποκειμενικά κριτήρια" ή αλλιώς QoE.

Το QoE (Quality of Experience) αναφέρεται στην τελική εμπειρία που έχει ο χρήστης και την γνώμη που αυτός τελικά αποκομίζει από την χρήση της υπηρεσίας. Η ITU ορίζει την QoE ως "την συνολική αποδοχή μιας εφαρμογής ή μιας υπηρεσίας όπως υποκειμενικά και μόνο την αξιολόγησε ο τελικός χρήστης", η ποιότητα εμπειρίας περιλαμβάνει ακόμα το σύνολο των ενδεχόμενων επιρροών όπως τον πελάτη, την τερματική συσκευή, το δίκτυο, τις παρεχόμενες υπηρεσίες και γενικότερα όλη την υποδομή. Αναφέρεται επίσης ότι η συνολική αποδοχή

μπορεί να επηρεαστεί και από τις προσδοκίες του χρήστη όπως και του περιβάλλοντος που η υπηρεσία λειτουργεί.

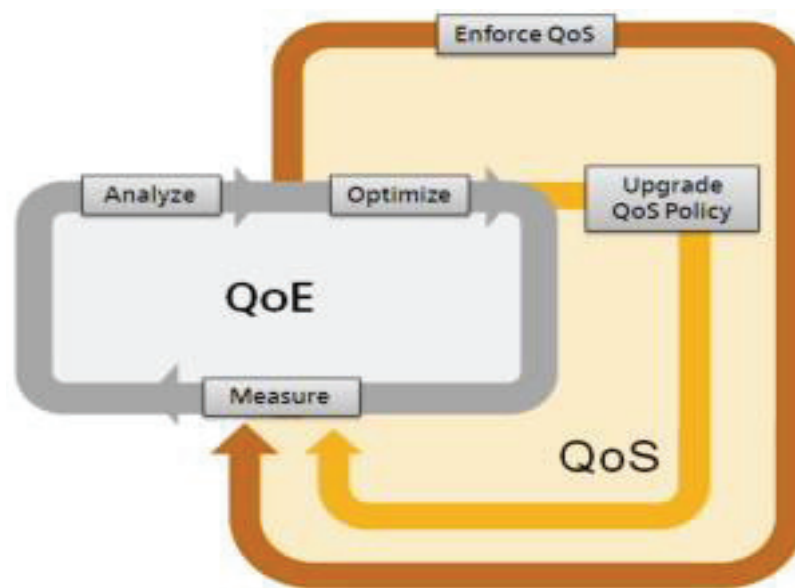
Παράλληλα με το QoE υπάρχει και ο όρος QoS (Quality of Service) ο οποίος χρησιμοποιείται από τους παρόχους, τους διαχειριστές δικτύων και τους κατασκευαστές εξοπλισμού σαν κοινός παρανομαστής δείκτη ποιότητας του υλικού.

Το QoS (Quality of Service) όπως ορίζεται από την ITU στο πρότυπο ITU X.902, η ποιότητα υπηρεσίας είναι ένα σύνολο απαιτήσεων ποιότητας επί της συνολικής συμπεριφοράς ενός ή παραπάνω αντικειμένων. Η ποιότητα υπηρεσίας αφορά σε χαρακτηριστικά της σύνδεσης όπως ο ρυθμός μετάδοσης πληροφορίας, καθυστέρησης του συστήματος κ.λπ.

Το ETSI ορίζει τις απαιτήσεις της ποιότητας υπηρεσίας στο πρότυπο του ETSI TS 102 250 και οι οποίες είναι:

- (Network Availability) Διαθεσιμότητα δικτύου: Είναι η δυνατότητα οι παρεχόμενες υπηρεσίες να διατίθενται στο χρήστη.
- (Network Accessibility) Προσβασιμότητα δικτύου: Είναι η δυνατότητα ο χρήστης να εγγράφει με επιτυχία στο δημόσιο δίκτυο.
- (Service Accessibility) Προσβασιμότητα υπηρεσίας: Είναι η δυνατότητα σε περίπτωση επιθυμίας του χρήστη να χρησιμοποιήσει μια υπηρεσία να αποκτήσει πρόσβαση από τον πάροχο το συντομότερο δυνατό.
- (Service Integrity) Ακεραιότητα υπηρεσίας: Περιγράφει την ποιότητα της υπηρεσίας κατά την λειτουργίας της.
- (Service Retainability) Διατηρησιμότητα υπηρεσίας: Αναφέρει τον τερματισμό των υπηρεσιών από τον χρήστη με τη θέληση του ή όχι.

Η διαφορά λοιπόν των όρων είναι ότι το μεν QoS υπολογίζει τεχνικές παραμέτρους εστιάζοντας στο αποτέλεσμα και την απόδοση της εκάστοτε υπηρεσίας ενώ το QoE συνυπολογίζοντας το QoS εστιάζει στην τελική γνώμη του χρήστη βασιζόμενο στην αντιληφθείσα ποιότητα του μέσου και την τελική ικανοποίηση του από την χρήση της εκάστοτε υπηρεσίας.<sup>1</sup>



Εικόνα 2

Ο όρος QoE είναι σχετικά νέος και συνεχώς εξελίσσεται. Παρόμοιοι με αυτόν όροι που χρησιμοποιούνται ώστε να προσεγγίσουν όσο το δυνατόν καλύτερα τον δείκτη της ποιότητας είναι:

- QoS (Quality of Service) - Ποιότητα Υπηρεσίας
- QoSE (Quality of Service Experienced) - Ποιότητα Εμπειρίας
- QoP (Quality of Perception) - Αντιλαμβανόμενη Ποιότητα

<sup>1</sup> Βλέπε εικόνα 2

Ο όρος QoSE όπως τον δίνει η ίδια η Διεθνής Ένωση Τηλεπικοινωνιών στο πρότυπο ITU-T Rec.E800 «είναι μια δήλωση που εκφράζει το επίπεδο της ποιότητας που πιστεύουν ότι βίωσαν οι χρήστες». Αφορά λοιπόν την ποιότητα της παρεχόμενης υπηρεσίας όπως την αντιλήφθηκε ο χρήστης σε επίπεδο ποιότητας άλλα και ποσότητας.

Ο όρος QoP αναφέρεται κυρίως στο αν γίνεται αντιληπτή μια αλλαγή στην ποιότητα της παρεχόμενης υπηρεσίας και αν αυτή η ποιότητα γίνεται αποδεκτή. Τυπικά η αντίληψη της ποιότητας αυτής από τον χρήστη μετριέται με την υποκειμενική κλίμακα αξιολόγησης και αναφέρεται σαν "Mean Opinion Score" ή MOS όπου θα αναλυθεί παρακάτω.

Οι παράγοντες που επιδρούν στην τελική αξιολόγηση της παρεχόμενης υπηρεσίας μπορούν να χωριστούν λοιπόν σε δυο βασικές κατηγορίες, τους ανθρωπογενείς και τους σχετικούς με τεχνικές παραμέτρους παράγοντες:

- Ανθρωπογενείς παράγοντες είναι: Προσδοκίες, Ηλικία ομιλητή και ακροατή, Φύλο ομιλητή και ακροατή, Οικειότητα μεταξύ ομιλητή και ακροατή, Μητρική Γλώσσα, κλπ.
- Τεχνικοί Παράγοντες QoS: Αλγοριθμος κωδικοποίησης φωνής, Συμπύεση φωνής, Κατάργηση περιόδων σιωπής, Εξουδετέρωση ήχους, Επανακωδικοποίηση φωνής, Καθυστέρηση δικτύου, Διακύμανση καθυστέρησης, Απώλεια πακέτων, Διαθεσιμότητα εύρους ζώνης δικτύου, κλπ.

### **2.1.2 Μέθοδοι αξιολόγησης φωνής**

Οι μέθοδοι αξιολόγησης της ομιλίας χωρίζονται σε δυο βασικές κατηγορίες, τις υποκειμενικές μεθόδους αξιολόγησης και τις αντικειμενικές μεθόδους αξιολόγησης.



- Οι υποκειμενικές μέθοδοι αξιολόγησης βασίζονται στους ανθρώπους οι οποίοι μέσα σε συγκεκριμένα τεχνικά πλαίσια που ορίζονται από τις προδιαγραφές της ITU προσπαθούν να αξιολογήσουν ποιοτικά τις παρεχόμενες υπηρεσίες. Ο βαθμός αξιολόγησης της ποιότητας των συγκεκριμένων μεθόδων θεωρείται ακόμα και σήμερα ο πιο αξιόπιστος διότι είναι ότι πιο αντιπροσωπευτικό στην πραγματικότητα. Ο βαθμός αυτός ονομάζεται MOS (Mean Opinion Score) ή Μέση γνωμοβαθμολογία.
- Οι αντικειμενικές μέθοδοι αντίθετα βασίζονται σε αλγόριθμους που προσπαθούν να εκτιμήσουν ποιοτικά τις παρεχόμενες υπηρεσίες όπως θα επιχειρούσε να το κάνει ο ανθρώπινος παράγοντας. Τα αποτελέσματα των αλγόριθμων αυτών προσπαθούν ουσιαστικά να προσεγγίσουν όσο το δυνατό πιο καλά το βαθμό αξιολόγησης της ποιότητας (MOS) που θα προέκυπτε από υποκειμενικές μεθόδους αξιολόγησης. Οι αντικειμενικές μέθοδοι χωρίζονται σε δύο βασικές κατηγορίες, τις παρεμβατικές αντικειμενικές μεθόδους αξιολόγησης της ομιλίας και τις μη παρεμβατικές αντικειμενικές μεθόδους αξιολόγησης της ομιλίας. Οι παραπάνω μέθοδοι αντικειμενικής αξιολόγησης θα παρουσιαστούν αναλυτικότερα στο κεφάλαιο 4.



Εικόνες 3 & 4



## ΚΕΦΑΛΑΙΟ 3

### ΥΠΟΚΕΙΜΕΝΙΚΗ ΜΕΘΟΔΟΣ ΑΞΙΟΛΟΓΗΣΗΣ

#### 3.1 MOS (Mean Opinion Score)

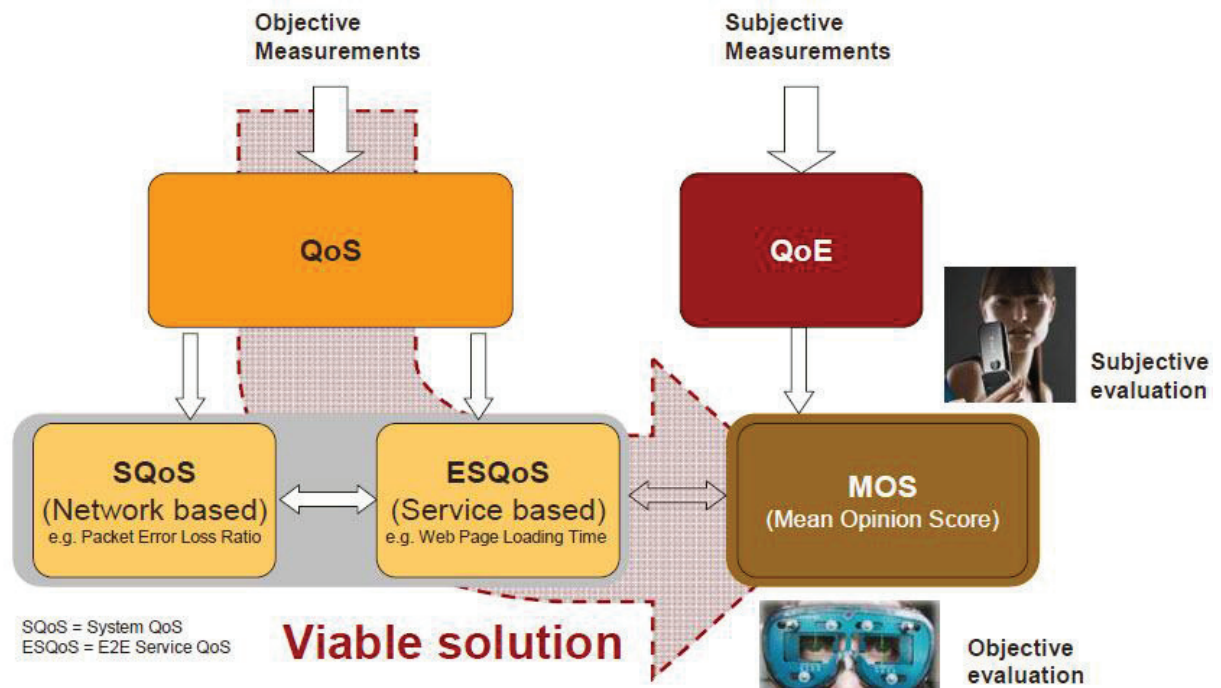
Το 1996 συστάθηκε από την Διεθνή Ένωση Τηλεπικοινωνιών το πρότυπο P.800 το οποίο περιγράφει τις μεθόδους υποκειμενικής αξιολόγησης που θεωρούνται ενδεδειγμένες για να καθοριστεί το πόσο ικανοποιητικά λειτουργούν τηλεφωνικές συνδέσεις, αλγόριθμοι, συσκευές ή και ολόκληρα τηλεπικοινωνιακά δίκτυα. Η μέση γνωμοβαθμολογία υπολογίζεται από τον μέσο όρο των βαθμολογιών των συμμετεχόντων οι οποίοι εκφράζουν την γνώμη τους για την επίδοση του τηλεφωνικού συστήματος μέσω συγκεκριμένων τιμών βαθμολόγησης σε προκαθορισμένη κλίμακα. Οι αξιολογήσεις αυτές του τηλεφωνικού συστήματος μπορούν να αφορούν είτε συνομιλίες είτε ακροάσεις ηχογραφημένου προφορικού υλικού.

Η MOS χρησιμοποιείται σαν γενική εφαρμογή αξιολόγησης χωρίς να υπολογίζει συγκεκριμένα παράγοντες υποβάθμισης όπως είναι: απώλειες, θόρυβος κυκλώματος, σφάλματα μετάδοσης, θόρυβος περιβάλλοντος, πλάγιος τόνος, ηχώ ομιλητή, μη γραμμική παραμόρφωση, χρόνος διάδοσης, επιβλαβείς επιπτώσεις συσκευών φωνής, παραμορφώσεις που προκύπτουν από την μεταγωγή πακέτου, χρονομεταβλητές υποβαθμίσεις του διαύλου επικοινωνίας κλπ. Στην ουσία συνυπολογίζοντας την ποιότητα εμπειρίας η οποία προέρχεται όμως από την ποιότητα υπηρεσίας ο συμμετέχων στην αξιολόγηση βαθμολογεί το σύστημα το οποίο είναι υπό δοκιμή.<sup>2</sup>

---

<sup>2</sup> Βλέπε Σχήμα 1

## QoS and QoE parameters – Mapping Model



Σχήμα 1: QoS and QoE parameters

Τα πειράματα της υποκειμενικής αξιολόγησης βάσει του προτύπου ITU-T P.800 χωρίζονται στις εξής κατηγορίες:

- Δοκιμές αξιολόγησης συνομιλίας (Conversation Opinion Tests)
- Δοκιμές αξιολόγησης ακρόασης (Listening Opinion Tests)
- Δοκιμές δημοσκοπικής έρευνας (Interview & Survey Tests)

Η προκαθορισμένη κλίμακα γνωμοβαθμολόγησης (MOS) ορίζεται από την ITU σε κλίμακα των πέντε μονάδων ξεκινώντας από το 1 που είναι κακή ποιότητα ομιλίας και φτάνοντας στο 5 που είναι εξαιρετική ποιότητα ομιλίας. Ο πίνακας αναλυτικά όπως τον ορίζει η ITU είναι:

Βαθμός MOS (Mean Opinion Score)	Ποιότητα Ομιλίας	Επίπεδο Υποβάθμισης
5	Excellent	Ανεπαίσθητη
4	Good	Αισθητή αλλά όχι ενοχλητική
3	Fair	Ελαφρώς ενοχλητική
2	Poor	Ενοχλητική
1	Bad	Πολύ Ενοχλητική

**Πίνακας 4: Κλίμακα γνωμοβαθμολόγησης**

### 3.1.1 Δοκιμές Αξιολόγησης συνομιλίας

Το συγκεκριμένο τεστ αξιολόγησης της συνομιλίας πραγματοποιείται σε εργαστηριακό περιβάλλον που έχει σαν σκοπό την όσο πιο δυνατόν καλύτερη αναπαράσταση των πραγματικών συνθηκών που βιώνουν οι χρήστες της τηλεφωνίας. Για να επιτευχθεί αυτό πρέπει να γίνει σωστή επιλογή των υποκειμένων (άτομα που συμμετέχουν στο πείραμα), τις συνθήκες του κυκλώματος (προσομοίωση συνθηκών υποβάθμισης) και να διαχειριστεί η όλη διαδικασία με τον τρόπο που πρέπει ώστε να εξαχθεί το επιθυμητό αποτέλεσμα. Είναι επίσης πολύ σημαντικό οι συνθήκες προσομοίωσης των δοκιμών να έχουν υλοποιηθεί βάση των προδιαγραφών που η ITU ορίζει. Έτσι λοιπόν στο συγκεκριμένο τεστ πρέπει οι συνθήκες προσομοίωσης να είναι σωστά υλοποιημένες βάση της ITU που της καθορίζει, να έχουν γίνει μετρήσεις με ακρίβεια πριν και μετά από κάθε πείραμα, ο τεχνικός εξοπλισμός να παρέχει την δυνατότητα πλήρους προσομοίωσης της κλήσης όπως η πληκτρολόγηση του

τηλεφωνικού αριθμού ή το χτύπημα του ήχου κλήσης της συσκευής στο τέλος της γραμμής και τα αρχεία καταγραφής όλων των αποτελεσμάτων από το πείραμα να κρατούνται πιστά.

Συγκεκριμένα οι προδιαγραφές του πειράματος έχουν ως εξής:

Τα δυο άτομα κάθονται σε ξεχωριστά ηχομονωμένα δωμάτια σχετικά κοντά στο σημείο που το πείραμα ελέγχεται. Ο χώρος του δωματίου δεν πρέπει να είναι μικρότερος από 20 τετραγωνικά μέτρα και με χρόνο αντήχησης λιγότερο από 500ms. Οι εσωτερικές διαστάσεις του δωματίου πρέπει να είναι τέτοιες ώστε οι αντηχήσεις να είναι το λιγότερο δυνατό, μια τυπική αναλογία δωματίου είναι 5:04:03. Η φυσική κατασκευή των δωματίων πρέπει να είναι τέτοια ώστε να εξασθενεί οποιοσδήποτε εξωτερικός θόρυβος εντός του χώρου. Επίσης ο χώρος θα πρέπει να είναι φυσικά διακοσμημένος και εξοπλισμένος προσομοιάζοντας όσο το δυνατόν καλύτερα ένα τυπικό χώρο πραγματικού περιβάλλοντος.



**Εικόνα 5**

Υπάρχει ένας μεγάλος αριθμός μεθόδων κατάλληλων για χρήση στο σχεδιασμό των πειραμάτων όπως π.χ. Latin Squares, Youden Squares, Balanced Incomplete Blocks και Randomization with Replication. Αυτές οι μέθοδοι ανάλογα με το τι θέλει ο διεξάγων το πείραμα, θα αποφασίσει ποια μέθοδο θα ακολουθήσει λαμβάνοντας υπόψη του τον αριθμό των δοκιμών που σκοπεύει να διεξάγει, την ακρίβεια των αποτελεσμάτων που θέλει και την ικανότητα να βγάλει ασφαλή συμπεράσματα από τις μετρήσεις που τελικά θα έχει στην κατοχή του. Πρέπει να γίνει κάθε δυνατή προσπάθεια ώστε να εξασφαλιστεί ότι οι συνομιλίες που διεξάγονται είναι προς όφελος του πειράματος και τα άτομα που συμμετέχουν σε αυτές έχουν την δυνατότητα της πλήρους εκμετάλλευσης των μεθόδων εκτομής και λήψης που το πείραμα παρέχει για το σκοπό αυτό. Ο γενικός κανόνας είναι ότι κάθε συζήτηση θα πρέπει να έχει ένα φυσιολογικό ξεκίνημα συνομιλίας και ένα φυσιολογικό τέλος αυτής. Σε καμία περίπτωση εκτός και αν αυτό είναι απαραίτητως αναγκαίο δεν πρέπει η συνομιλία να τερματίζεται στο μέσον της διάρκειας του πειράματος. Τα άτομα που λαμβάνουν μέρος στο πείραμα επιλέγονται τυχαία από τον πληθυσμό που χρησιμοποιεί το τηλέφωνο με τις εξής όμως προϋποθέσεις:

- i) Δεν έχουν καμία συμμετοχή σε όλες τις εργασίες που λαμβάνει χώρα το πείραμα όπως την αξιολόγηση της απόδοσης των τηλεφωνικών κυκλωμάτων ή την κωδικοποίηση της ομιλίας κ.λπ.
- ii) Δεν έχουν συμμετάσχει σε οποιαδήποτε έρευνα υποκειμενικής αξιολόγησης τους προηγούμενους έξι μήνες και συγκεκριμένα για πείραμα αξιολόγησης συνομιλίας το όριο μη συμμετοχής είναι τουλάχιστον ένα χρόνο πριν.

Εάν ο διαθέσιμος πληθυσμός περιορίζεται κατά το ελάχιστο χωρίς το επαρκές για τις απαιτήσεις του πειράματος δείγμα τότε τα αποτελέσματα που θα εξαχθούν από αυτό θα πρέπει να το λάβουν και αυτό υπ' όψιν. Δεν λαμβάνονται συγκεκριμένα μέτρα για να ισορροπιστεί ο αριθμός των γυναικών και των αντρών που λαμβάνουν μέρος στο πείραμα

εκτός αν αυτό απαιτείται από τον σχεδιασμό του συγκεκριμένου πειράματος. Στην αξιολόγηση του πειράματος χρησιμοποιείται η ευρέως διαδεδομένη κλίμακα κρίσης των πέντε σημείων . Η διάταξη της κλίμακας αυτής για το πείραμα αξιολόγησης της συνομιλίας, με πολύ μικρές παραλλαγές από γλώσσα σε γλώσσα βάση της πρωτότυπης όπως έχει στο αγγλικό κείμενο, έχει ως εξής :

- i. Excellent - 5 - Άριστη
- ii. Good - 4 - Καλη
- iii. Fair - 3 - Μετρια
- iv. Poor - 2 - Φτωχη
- v. Bad - 1 - Κακη

Στο τέλος κάθε συνομιλίας το άτομο που λαμβάνει μέρος στο πείραμα την αξιολογεί βάση της παραπάνω κλίμακας αξιολόγησης βάζοντας τον αριθμό που αντιστοιχεί στην εμπειρία που βίωσε. Ο μέσος όρος λοιπόν που προκύπτει στο τέλος του πειράματος από όλα τα άτομα που συμμετέχουν και όλα τα τεστ που διενεργήθηκαν ονομάζεται MOS.

### **3.1.2 Δοκιμές Αξιολόγησης Ακρόασης**

Προκειμένου να εξαλειφθούν τα ανεπιθύμητα προβλήματα που παρουσιάζονται στην μεταβλητότητα του λόγου σε πειράματα αξιολόγησης συνομιλίας, προσδιορίστηκε ένα πείραμα άλλης μορφής υποκειμενικής αξιολόγησης, αυτό της αξιολόγησης ακρόασης.

Το πείραμα αυτό ουσιαστικά διαφοροποιείται στα έτοιμα δείγματα ομιλίας που χρησιμοποιούνται σε σχέση με το φυσικό "ζωντανό" δείγμα ομιλίας της προηγούμενης μεθόδου. Τα ηχητικά σήματα - δείγματα έχουν εξαρχής κατασκευαστεί σε τέτοια μορφή που έχουν όλες εκείνες τις τυποποιημένες ιδιότητες που το πείραμα χρειάζεται για το σκοπό που



διενεργείται για να εξάγει όσο το δυνατόν ασφαλέστερα συμπεράσματα. Για να επιτευχθεί αυτό χρειάζονται να τηρηθούν οι εξής προϋποθέσεις:

i) Περιβάλλον καταγραφής ηχητικού σήματος

Ο ομιλητής θα πρέπει να κάθεται σε ήσυχο δωμάτιο με εμβαδόν μεταξύ 30 m<sup>3</sup> και 120 m<sup>3</sup> και με χρόνο αντήχησης μικρότερο από 500 ms με επιθυμητό χρόνο μεταξύ 200-300 ms. Επίσης το επίπεδο θορύβου μέσα στο δωμάτιο θα πρέπει να είναι κάτω από 30 dBA.

ii) Σύστημα αποστολής ηχητικού σήματος

Το σύστημα αποστολής του ηχητικού σήματος μπορεί να είναι μια απλή τηλεφωνική σύνδεση ή ενδιάμεσο σύστημα αναφοράς (IRS). Αυτό που έχει σημασία είναι το σύστημα να ρυθμιστεί σύμφωνα με το προδιαγραφές του πειράματος και η όποια χαρακτηριστική ευαισθησία συχνότητας παρατηρηθεί να αναφερθεί άμεσα ώστε να αξιολογηθεί ανάλογα. Η χαρακτηριστική ευαισθησία της σύνδεσης μετριέται τουλάχιστον δυο φορές στην αρχή και το τέλος κάθε πειράματος. Κάθε σημαντική διαφορά στις δυο αυτές μετρήσεις πρέπει να εκτιμηθεί ανάλογα από τον επιβλέπων του πειράματος και να κριθεί ανάλογα για τυχόν μη εγκυρότητα των αποτελεσμάτων αυτού.

iii) Σύστημα καταγραφής ηχητικού σήματος

Το σύστημα καταγραφής πρέπει να είναι υψηλού επιπέδου και μπορεί να είναι :

- Ένα συμβατικό μαγνητόφωνο κασέτας με υψηλής ποιότητας κασέτα καταγραφής με χαμηλό αποτύπωμα καταγραφής και χαμηλό επίπεδο θορύβου.
- Ένας ψηφιακός επεξεργαστής ήχου δυο καναλιών με υψηλής ποιότητας μηχανήμα καταγραφής βιντεοκασέτας ή κασέτας ψηφιακού ήχου.
- Ακόμα μπορεί να χρησιμοποιηθεί ελεγχόμενο από υπολογιστή ψηφιακό σύστημα αποθήκευσης.

Το τελευταίο σύστημα θεωρείται καλύτερο και πιο ευέλικτο καθαρά για πρακτικούς λόγους παραταύτα μερικές φορές αποτελεί μονόδρομος η επιλογή ενός από τα άλλα δυο συστήματα λόγω συγκεκριμένων απαιτήσεων του εκάστοτε πειράματος.

#### iv) Περιεχόμενο ομιλίας

Το περιεχόμενο της ομιλίας θα πρέπει να αποτελείται από απλές, ουσιαστικές και σύντομες φράσεις που έχουν επιλεγεί τυχαία και είναι εύκολα κατανοητές π.χ. από την τρέχουσα βιβλιογραφία ή τις εφημερίδες. Αυτές οι φράσεις θα πρέπει να είναι δομημένα σε λίστες σε τυχαία σειρά ώστε να μην υπάρχει καμία προφανής σύνδεση νοήματος μεταξύ τους. Πολύ μικρές φράσεις ομιλίας ή πολύ μεγάλες θα πρέπει να αποφεύγονται διότι ο σκοπός είναι οποιοδήποτε κείμενο εκφωνηθεί από τον ομιλητή που συμμετέχει στο πείραμα να γίνεται σε μια περίοδο χρόνου μεταξύ δυο το πολύ τριών δευτερολέπτων. Παραδείγματα αυτών των φράσεων παρουσιάζονται στον πίνακα 5.

Ο διεξάγων το πείραμα θα πρέπει να αποφασίσει πόσα κείμενα - φράσεις χρειάζονται σε κάθε ομάδα ώστε να αποτελούν ένα δείγμα ηχητικού σήματος - ομιλίας. Προτείνεται τουλάχιστον δυο φράσεις και το περισσότερο πέντε. Το κενό χρονικό διάστημα μεταξύ των φράσεων κατά το οποίο μπορεί να ακουστεί θόρυβος του κυκλώματος είναι πολύ σημαντικό. Προτείνεται να γίνεται εγγραφή όσο τον δυνατόν μεγαλύτερων ομάδων καταγραφής ομιλίας αν και με μικρότερες ομάδες μπορεί το επιτευχθεί ένα πείραμα χρησιμοποιώντας την επανάληψη ή την αντιγραφή των εγγραφών τους.

<b><i>Παραδείγματα από κείμενο ομιλίας</i></b>
<b>Θα πρέπει να είσαι πολύ ήσυχος</b>
<b>Δεν υπάρχει τίποτα να δεις εδώ</b>
<b>Κατασκευάζουν ξύλινα είδωλα</b>
<b>Χρειάζομαι ένα λεπτό με τον εξεταστή</b>
<b>Χρειάζεσαι καθόλου χρήματα;</b>

**Πίνακας 5**

v) Διαδικασία εγγραφής

Η ομιλία καταγράφεται από ένα μικρόφωνο το οποίο είναι τοποθετημένο μεταξύ 14 εκατοστών και 20 εκατοστών από τα χείλη του ομιλητή. Δυο ξεχωριστά συστήματα χρησιμοποιούνται ταυτόχρονα, το ένα για την καταγραφή της ευρυζωνικής ομιλίας σε ένα κανάλι και ένα για την καταγραφή της τηλεφωνικής ομιλίας στο ίδιο κανάλι. Συνήθως η μια από τις δυο καταγραφές είναι αρκετή αλλά σε κάποιες περιπτώσεις το πείραμα απαιτεί την χρήση και των δυο συστημάτων.

vi) Ομιλητές

Πρέπει να υπάρχουν τόσοι ομιλητές όσους ο σχεδιασμός του πειράματος απαιτεί. Είναι φυσικό ότι η εκφορά του λόγου πρέπει να είναι ομαλή χωρίς τραυλίσματα ή δραματικότητες. Θα συμμετέχουν άτομα και των δυο φύλων ώστε να βγει ένα ισορροπημένο αποτέλεσμα. Παραταύτα θα πρέπει να μετριοούνται ξεχωριστά τα αποτελέσματα τους και μόνο στην περίπτωση που δεν υπάρχουν σημαντικές αποκλίσεις τα αποτελέσματα αυτά θα αθροίζονται. Επίσης για να είναι απόλυτα σαφές ότι το αποτέλεσμα του πειράματος δεν επηρεάζεται δραματικά από χροιές και τόνους φωνής συγκεκριμένων ανδρών ή γυναικών καλό είναι να συμμετέχουν στο πείραμα πάνω από ένας άνδρας και πάνω από μια γυναίκα.

vii) Ακροατές

Τα άτομα που λαμβάνουν μέρος στο πείραμα σαν ακροατές επιλέγονται τυχαία από τον πληθυσμό που χρησιμοποιεί το τηλέφωνο με τις προϋποθέσεις<sup>3</sup> που ορίζει η ITU. Το σύστημα αξιολόγησης που χρησιμοποιούν είναι το MOS με την αντίστοιχη κλίμακά του από το 5 έως το 1 και από το εξαιρετικό έως το κακό αντίστοιχα.

---

<sup>3</sup> Βλέπε παράγραφο 3.1.1



**Εικόνα 6**

Η συνιστώμενη από την ITU μέθοδος για τις δοκιμές ακρόασης ομιλίας είναι η ACR (Absolute Category Rating) και η διαδικασία αυτής αναλύθηκε πιο πάνω. Παρατάτα υπάρχουν και άλλες μέθοδοι που χρησιμοποιούνται για πολύ συγκεκριμένους σκοπούς και όχι τόσο διαδεδομένα όπως η ACR. Αυτές είναι η DCR (Degradation Category Rating), η CCR (Comparison Category Rating), η Quintal-Response Detectability Test και η Threshold Method τις οποίες αναλυτικά αναφέρει η σύσταση της ITU P.800.

### **3.1.3 Δοκιμές Δημοσκοπικής Έρευνας**

Ένας άλλος τρόπος υποκειμενικής αξιολόγησης σε περίπτωση που υπάρχει ο απαιτούμενος χρόνος και οι προδιαγραφές του πειράματος τον απαιτούν είναι οι δοκιμές δημοσκοπικής έρευνας. Αφορά πειράματα στα οποία η ποιότητα μπορεί να προσδιοριστεί και να βαθμολογηθεί με δημοσκοπικές έρευνες σε χρήστες τηλεφωνικών υπηρεσιών και αφορά την ίδια την υπηρεσία που χρησιμοποιούν. Οι έρευνες αυτές εκτός από χρονοβόρες έχουν και

πολύ συγκεκριμένες προδιαγραφές χωρίς την δυνατότητα ενός μεγάλου εύρους αξιολόγησης υπηρεσιών, συσκευών ή και επί μέρους χαρακτηριστικών των υπό αξιολόγηση τηλεφωνικών συνδέσεων, οπότε και αξιοποιούνται πρωτίστως από εταιρίες παροχής τηλεφωνίας σαν άμεσες και πλέον ανέξοδες για τις ίδιες, με το όφελος της διαπροσωπικής επαφής με τον πελάτη και ικανοποίησης των αιτημάτων ή παραπόνων του. Οι τρόποι διεξαγωγής των δημοσκοπικών αυτών ερευνών ποικίλουν ανάλογα με τις ανάγκες και πολιτικές που έχει η εκάστοτε τηλεφωνική εταιρία, παραταύτα η ITU έχει δημοσιοποιήσει προδιαγραφές των ερευνών αυτών και στη σύσταση ITU-T Rec.P.82.



## **ΚΕΦΑΛΑΙΟ 4**

### **ΑΝΤΙΚΕΙΜΕΝΙΚΗ - ΑΛΓΟΡΙΘΜΙΚΗ ΜΕΘΟΔΟΣ ΑΞΙΟΛΟΓΗΣΗΣ**

#### **4.1 Παρεμβατικές αντικειμενικές μέθοδοι**

Οι παρεμβατικές αντικειμενικές μέθοδοι παρέχουν δείκτες ποιότητας ομιλίας συγκρίνοντας τα αρχικά δείγματα ομιλίας που αποστέλλονται και τα τελικά υποβαθμισμένα δείγματα που παραλαμβάνονται ως εκ τούτου απαιτούν πρόσβαση και στα δυο άκρα της επικοινωνίας, αυτό της μετάδοσης της ομιλίας και αυτό της λήψης της. Συγκρίνοντας λοιπόν τα δείγματα αναφοράς και τα υποβαθμισμένα δείγματα ομιλίας γίνεται πιο ακριβής εκτίμηση της υποκειμενικής αντίληψης της ποιότητας ομιλίας που λαμβάνεται από την τερματική συσκευή. Παραταύτα η ακριβής αυτή εκτίμηση της ποιότητας διεξάγεται υπό το κόστος δοκιμής των δειγμάτων που στέλνονται μέσα από ένα ταυτόχρονα υπό δοκιμή δίκτυο τηλεφωνίας. Είναι λογικό ότι η τηλεφωνική σύνδεση κάτω από την οποία πραγματοποιείται το πείραμα να είναι μη διαθέσιμη για τους υπόλοιπους χρήστες της τηλεφωνίας. Επίσης σε περιπτώσεις που το πείραμα πραγματοποιείται σε ώρες αιχμής είναι δυνατόν τα αποτελέσματα να παρουσιάζουν αρκετά χαμηλότερη βαθμολογία από το αναμενόμενο. Παρ' όλα αυτά επειδή οι αντικειμενικές μέθοδοι πραγματοποιούν μετρήσεις ποιότητας από άκρη σε άκρη για το τηλεφωνικό δίκτυο είναι πολύ χρήσιμες για τους παρόχους τηλεφωνικών δικτύων που χρειάζονται άμεση παρακολούθηση της παρεχόμενης ποιότητας που προσφέρουν συνολικά τα δίκτυά τους.

### 4.1.1 Επίδοση του αλγόριθμου PSQM

Το 1996 η Διεθνής Ένωση Τηλεπικοινωνιών (ITU) δημοσίευσε τη μέθοδο PSQM (Perceptual Speech Quality) που ορίζεται στο πρότυπο ITU-T P.861 για την αντικειμενική αξιολόγηση της ποιότητας μετάδοσης φωνής (300 – 3400 Hz). Η PSQM είναι μια αυτοματοποιημένη μέθοδος που αναπτύχθηκε για την μέτρηση της αντιλαμβανόμενης ποιότητας σε φωνή πραγματικού χρόνου η οποία επηρεάζεται από κωδικοποιητές συμπίεσης. Συγκρίνει ένα διαστρεβλωμένο δείγμα φωνής με το πρωτότυπο δείγμα και χρησιμοποιεί μια περίπλοκη αναλυτική διαδικασία για την εκτίμηση της διαφοράς όσον αφορά τους παράγοντες που επηρεάζουν τις αντιλήψεις των ακροατών. Η τελική βαθμολογία παραμόρφωσης αντιστοιχεί σε μεγάλο βαθμό με τη μέση βαθμολογία γνώμης (MOS) που προκύπτει από την χρησιμοποιούμενη ομάδα ακροατών με τη μέθοδο P.800.

Ο λόγος που χρησιμοποιούμε αυτό τον αλγόριθμο είναι διότι με το PSQM επιτρέπεται η αυτοματοποιημένη, βασισμένη σε δοκιμές προσομοίωσης, αξιολόγηση σαφήνειας του λόγου με αντικειμενικά κριτήρια και η αναμεταδιδόμενη ποιότητα φωνής. Διάφορα προϊόντα λογισμικού ή και υλικού έχουν αναπτυχθεί για να διευκολυνθεί αυτή η δοκιμή. Αυτό έχει σαν αποτέλεσμα σημαντική εξοικονόμηση κόστους και χρόνου σε σχέση με την παραδοσιακή πρακτική της χρησιμοποίησης μεγάλων ομάδων ανθρώπων που αξιολογούν υποκειμενικά σήματα φωνής και βαθμολογούν την ποιότητα της φωνής. Επιπλέον δίνει αντικειμενικά αποτελέσματα τα οποία είναι αξιόπιστα και μπορούν να αναδημοσιευτούν. Αυτό είναι πολύ σημαντικό για τους παρόχους τηλεφωνίας που θέλουν να διατηρήσουν τα υψηλά πρότυπα ποιότητας των υπηρεσιών τους.



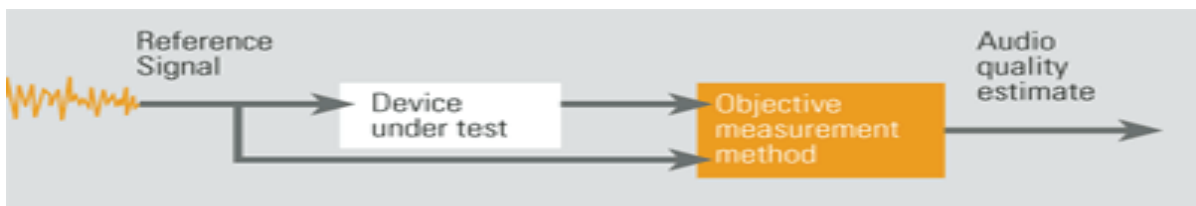
Ο αλγόριθμος που χρησιμοποιεί το PSQM αποτελείται από ένα ψυχοακουστικό<sup>4</sup> μαθηματικό μοντέλο (αντιληπτικών και γνωστικών λειτουργιών) αλγόριθμου για να αναλύσει τα προ και μετά μεταδιδόμενα σήματα φωνής, αποδίδοντας μια τιμή η οποία δηλώνει την μέτρηση της ποιότητας υποβαθμίσης του σήματος και κυμαίνεται από 0 (καμία υποβάθμιση) έως 6.5 (υψηλότερη υποβάθμιση). Η τιμή αυτή μπορεί να μεταφραστεί σε μια μέση βαθμολογία γνώμης (MOS), που είναι ένα αποδεκτό μέτρο της αντιληπτής ποιότητας των λαμβανόμενων μέσων σε μια αριθμητική κλίμακα που κυμαίνεται από το 1 έως το 5. Η τιμή 1 υποδεικνύει απαράδεκτη, κακή ποιότητα φωνής, ενώ η τιμή 5 υποδεικνύει υψηλή ποιότητα φωνής. Η ποιότητα της κωδικοποιημένης ομιλίας κρίνεται από τις διαφορές στην εσωτερική αναπαράσταση. Η διαφορά αυτή χρησιμοποιείται για τον υπολογισμό της διαταραχής του θορύβου ως συνάρτηση του χρόνου και της συχνότητας. Εκτός από την αντιληπτική μοντελοποίηση, ο PSQM χρησιμοποιεί γνωστική μοντελοποίηση, όπως η κλιμάκωση της έντασης και ασύμμετρη συγκάλυψη προκειμένου να πάρει υψηλές συσχετίσεις μεταξύ των υποκειμενικών και των αντικειμενικών μετρήσεων.

Ο αλγόριθμος PSQM παρέχει πιο αξιόπιστα και συνεπή αποτελέσματα MOS εάν χρησιμοποιούνται σύμφωνα με τις μεθόδους που προτείνει η ITU για την αντικειμενική και υποκειμενική εκτίμηση της ποιότητας (ITU-T P.800/P.830/P.861). Αυτές οι μέθοδοι περιλαμβάνουν ως σημείο αναφοράς τη χρήση σημάτων φωνής και των δυο φύλων (τόσο ανδρικών όσο και γυναικείων) σε ένα μέσο επίπεδο της τάξεως των -20dB. Το είδος της ομιλίας, το φύλο, η διάρκεια του σήματος μπορούν να έχουν μικρή επίπτωση στην PSQM τιμή ή την τιμή MOS σε αντίθεση με τον αριθμό των κλήσεων που πραγματοποιήθηκαν και διάφορων άλλων ρυθμίσεων διαμόρφωσης του περιβάλλοντος.

---

<sup>4</sup> Ψυχοακουστική (psychoacoustics) ονομάζεται η επιστήμη που έχει ως θέμα μελέτης τον υποκειμενικό τρόπο με τον οποίο το ανθρώπινο σύστημα ακοής αντιλαμβάνεται τους ήχους.

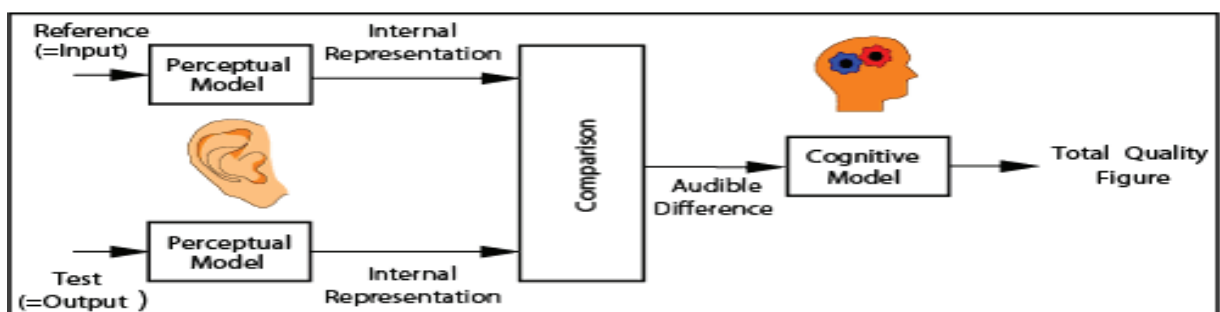
Η PSQM όμως ακριβώς επειδή ξεκίνησε να χρησιμοποιείται μόνο για την αξιολόγηση κωδικοποιητών ομιλίας και δεν είχε την δυνατότητα να υπολογίζει την μεταβλητή καθυστέρηση που υπήρχε, το φιλτράρισμα και τις τοπικές στρεβλώσεις, έγιναν αμέσως αντιληπτοί οι περιορισμοί αυτοί και για το λόγο αυτό αντικαταστάθηκε το 2000 από την Διεθνή Ένωση Τηλεπικοινωνιών με την μεθοδο P.862 γνωστή και ως PESQ η οποία βασίζεται σε ένα βελτιωμένο αλγόριθμο αντικειμενικής αξιολόγησης της ποιότητας φωνής.



Σχήμα 2

#### 4.1.2 Επίδοση του αλγόριθμου PEAQ

Μέθοδος PEAQ (Perceptual Evaluation Of Audio Quality) ονομάζεται μια ενεργητική αντικειμενική μέθοδος αξιολόγησης της ποιότητας του ήχου, η οποία στηρίζεται σε γενικά ψυχοακουστικές αρχές. Αναπτύχθηκε κατόπιν συνεργασίας πολλών ερευνητικών κέντρων από όλο τον κόσμο και ορίστηκε από την Διεθνή Ένωση Τηλεπικοινωνιών το 1999 με το πρότυπο ITU-R BS.1387



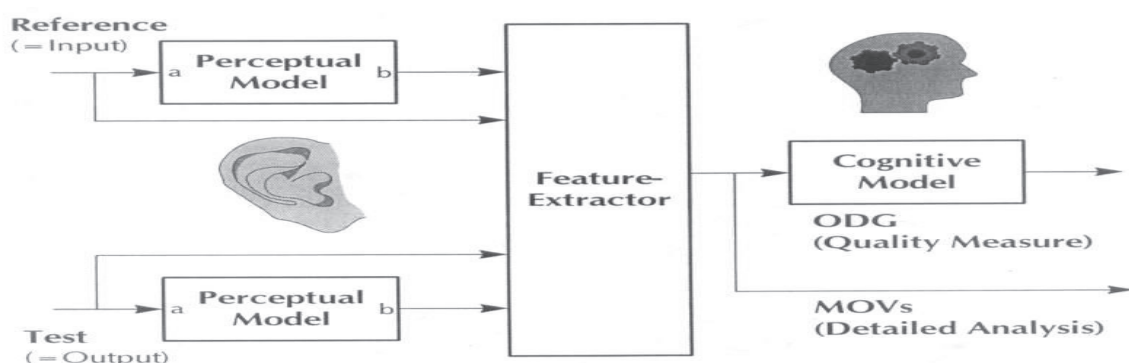
Εικόνα 7

Η PEAQ βασίζεται στο μοντέλο που παρουσιάζεται στην Εικόνα 7. Ουσιαστικά αυτό που κάνει είναι να συγκρίνει το αρχικό πρωτότυπο σήμα με το επεξεργασμένο ίδιο σήμα. Εν συνεχεία τα δυο αυτά σήματα μετατρέπονται σε αναπαράσταση μεμβράνης όπως αυτή που βρίσκεται στον κοχλία του ανθρώπινου αυτιού και ύστερα οι διαφορές που προκύπτουν αναλύονται σαν συναρτήσεις της συχνότητας και του χρόνου από ένα γνωστικό μοντέλο. Το μοντέλο αυτό παρουσιάζει τα χαρακτηριστικά εκείνα που αφορούν την αντίληψη και χρησιμοποιούνται για τον υπολογισμό του μέτρου της ποιότητας. Όπως φαίνεται στην Εικόνα 8 το μοντέλο συνυπολογίζει και άλλες ενδιάμεσες μεταβλητές εξόδου (MOV- model output variables). Οι ενδιάμεσες αυτές παράμετροι που έχουν προεπιλεχθεί αντιστοιχούν στο σύνολο τους σε ένα βαθμό ποιότητας. Η βαθμολογία ποιότητας που χρησιμοποιεί η μέθοδος PEAQ είναι η 5-βαθμια κλίμακα υποβάθμισης που έχουμε ήδη προαναφέρει και ορίζεται από την ITU.

Ο βαθμός που προκύπτει ονομάζεται ODG (Objective Difference Grade) κατά αντιστοιχία του SDG (Subjective Difference Grade) που προκύπτει από τον εξής τύπο:

$$\text{SDG} = \text{Βαθμός σήματος υπό δοκιμή} - \text{Βαθμός σήματος αναφοράς}$$

Συνοψίζοντας η μέθοδος PEAQ προσομοιάζοντας τον τρόπο λειτουργίας του ανθρώπινου αυτιού ενσωματώνει πολλαπλά μοντέλα μεταβλητών εξόδου (MOVs) και συνδυάζοντας τις διαφορετικές τιμές αυτών μεταξύ τους καταλήγει σε ένα βαθμό τύπου MOS προσομοιάζοντας την υποκειμενική αξιολόγηση ποιότητας ομιλίας.



Εικόνα 8

Η μέθοδος PEAQ προστατεύεται από διάφορα διπλώματα ευρεσιτεχνίας και είναι διαθέσιμη με άδεια μαζί με τον αρχικό κώδικά της για εμπορικές εφαρμογές σύμφωνα πάντα με τους όρους της Διεθνούς Ένωσης Τηλεπικοινωνιών. Για εκπαιδευτική χρήση υπάρχει η δωρεάν πλατφόρμα που ονομάζεται Peaqb, ένα πρόγραμμα το οποίο πραγματοποιεί ακριβώς τις ίδιες λειτουργίες απλά πιο περιορισμένα καθώς δεν είναι επικυρωμένο με τα στοιχεία της ITU. Ακόμη μη επικυρωμένη εφαρμογή για εκπαιδευτική χρήση του PEAQ είναι το PQeval Audio το οποίο είναι διαθέσιμο από το εργαστήριο TSP του Πανεπιστημίου McGill.

### **4.1.3 Επίδοση του αλγόριθμου PESQ**

Η PESQ (Perceptual Evaluation of Speech Quality) ή αλλιώς «Αντιληπτική αποτίμηση της ποιότητας ομιλίας» είναι μια αντικειμενική μέθοδος εκτίμησης της ποιότητας ομιλίας από άκρο σε άκρο για δίκτυα τηλεφωνίας στενής ζώνης (narrow-band).

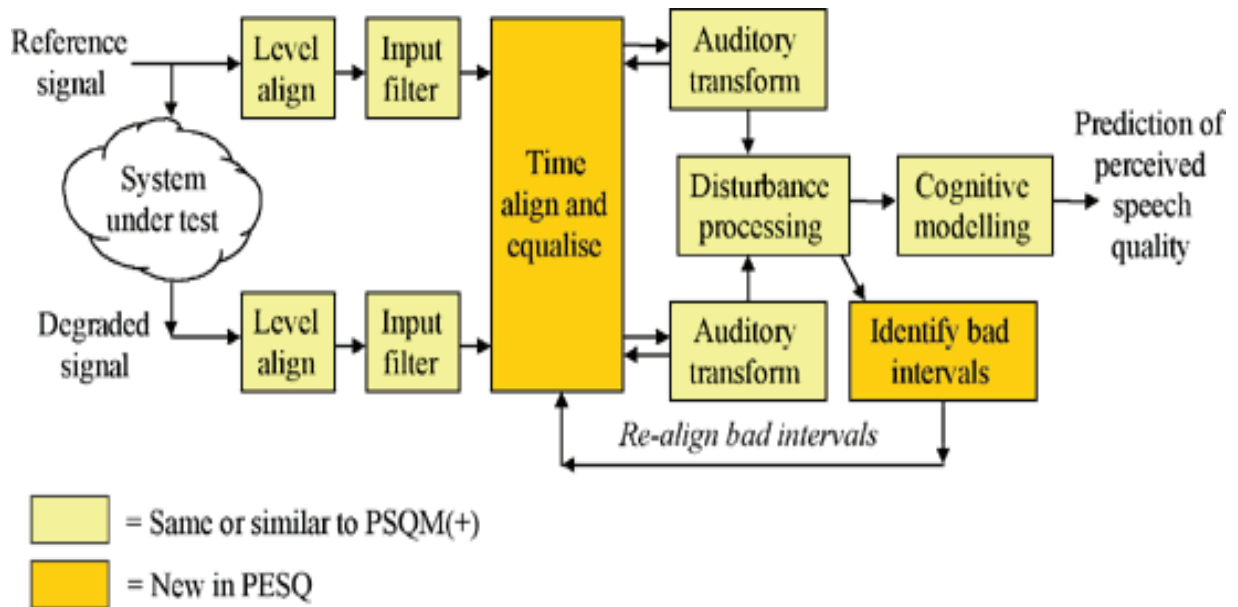
Η ανάπτυξη της πρόεκυψε μεν από την ανάγκη αντικειμενικών μετρήσεων σε επίπεδο αξιοπιστίας, όπως της MOS χωρίς το μεγάλο κόστος και την χρονοβόρα διαδικασία που αυτή απαιτεί, αλλά και για να διορθώσει δε όλους εκείνους τους περιορισμούς<sup>5</sup> που η μέθοδος PSQM είχε εξαρχής. Αποτελείται από μια ολόκληρη οικογένεια προτύπων και βασίζεται σε ένα πολύ πιο βελτιωμένο αλγόριθμο αντικειμενικής αξιολόγησης της ποιότητας ομιλίας λαμβάνοντας υπ' όψιν πλέον το φιλτράρισμα, την μεταβλητή καθυστέρηση, την παραμόρφωση της κωδικοποίησης και τα σφάλματα καναλιού<sup>6</sup>. Η δημοσίευση του έγινε το 2001 από την ITU-T στο πρότυπο P.862 και αρχικά αφορούσε για επικοινωνία στενής ζώνης

---

<sup>5</sup> Βλέπε παράγραφο 4.1.1

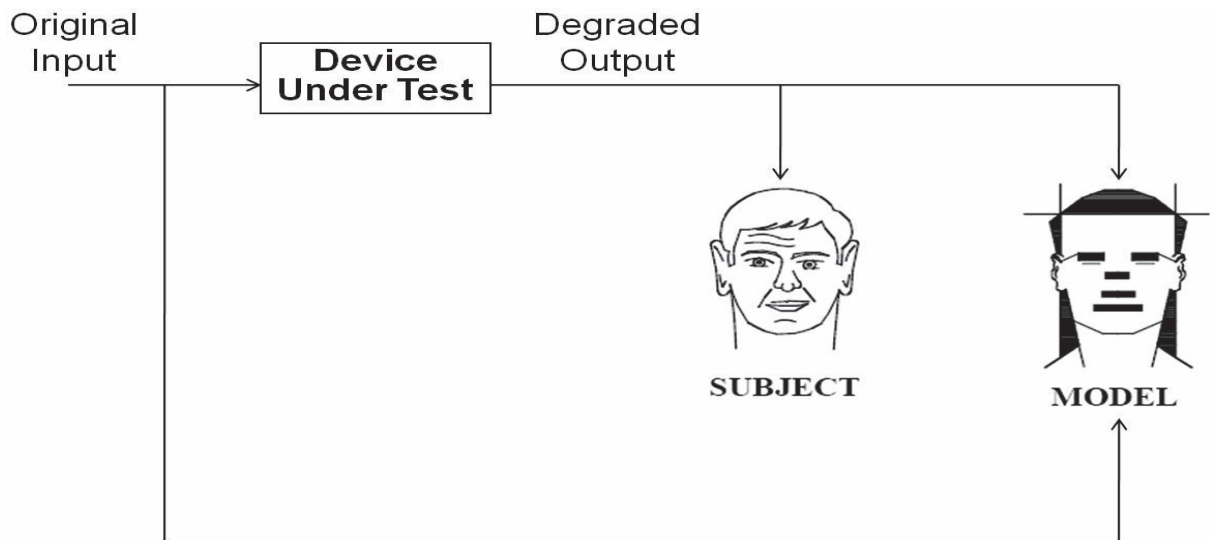
<sup>6</sup> Βλέπε Σχήμα 3

(narrow-band) και κωδικοποιητών φωνής αλλά στην πορεία επεκτάθηκε και σε ευρυζωνικά συστήματα (wide-band).



Σχήμα 3

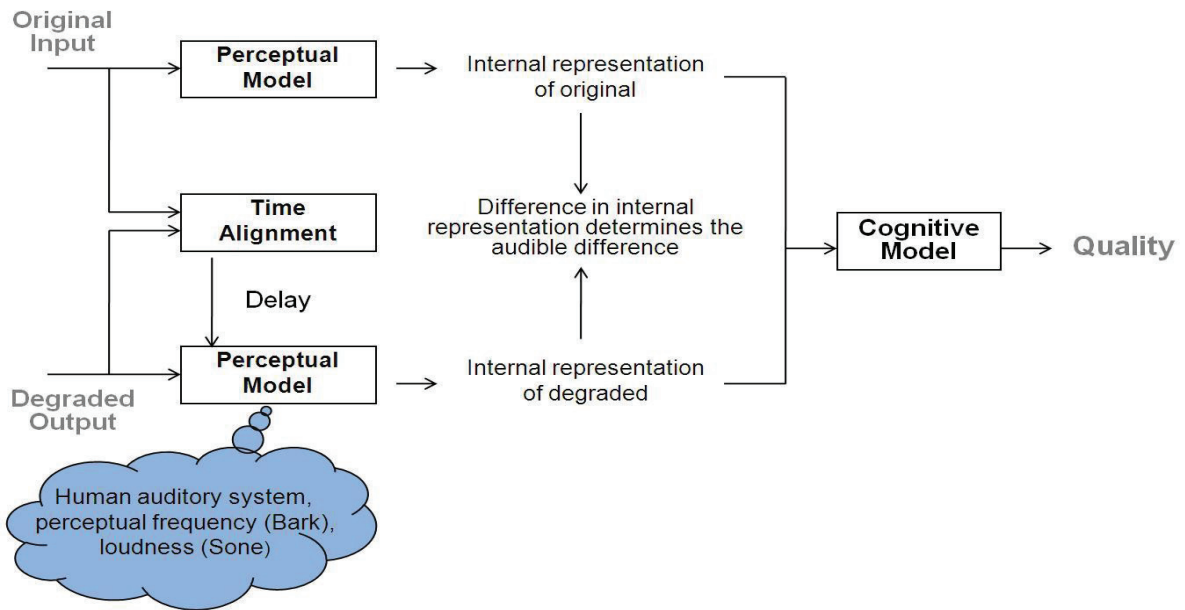
Η μέθοδος αυτή είναι η πλέον διαδεδομένη στην εποχή μας και αποτελεί ένα παγκόσμιο πρότυπο που εφαρμόζεται στην βιομηχανία κατασκευαστών τηλεφώνων και άλλων περιφερειών τηλεφωνίας όπως και κατασκευαστών εξοπλισμού δικτύων. Πρόκειται για το αποτέλεσμα πολλών ετών έρευνας και χρησιμοποιείται όχι μόνο για κωδικοποιητές ομιλίας αλλά και για μετρήσεις από άκρη σε άκρη. Ο λόγος που έχει επικρατήσει απόλυτα είναι διότι εκτος από τον έλεγχο της ποιότητας φωνής διορθώνει όλες εκείνες τις ατέλειες των άλλων προτύπων κρατώντας ταυτόχρονα όμως το στοιχείο που την κάνει τόσο αποδοτική, το στοιχείο της ανθρώπινης αντίληψης που δίνει όσο το δυνατό κοντινότερα αποτελέσματα με τη μέθοδο MOS.



**Εικόνα 9**

Στη πραγματικότητα λοιπόν αυτό που κάνει είναι να συγκρίνει το αρχικό σήμα ομιλίας, κρατώντας το σαν σημείο αναφοράς, με ένα υποβαθμισμένο σήμα το οποίο προέρχεται από το αρχικό αλλά έχει περάσει μέσα από ένα δίκτυο το οποίο είναι προς μέτρηση και σαν αποτέλεσμα προσπαθεί να αποδώσει την διαφορά μεταξύ του ιδανικού μοντέλου και του υποβαθμισμένου σήματος. Επίσης αναπαριστά το αποτέλεσμα σε μορφή υποκειμενικής αξιολόγησης ακρόασης. Πριν κάνει όμως την σύγκριση ο αλγόριθμος χρησιμοποιεί ένα ψυχοακουστικό και ένα γνωστικό μοντέλο. Το ψυχοακουστικό μοντέλο υπολογίζει την ψυχοσωματική αναπαράσταση των σημάτων φωνής και εφαρμόζεται τόσο στο αρχικό σήμα ομιλίας όσο και στο υποβαθμισμένο. Στην ουσία μετασχηματίζει το σήμα ομιλίας από το φυσικό του πεδίο στο ψυχοσωματικό το οποίο αντιστοιχεί στην ανθρώπινη αντίληψη.<sup>7</sup>

<sup>7</sup> Βλέπε εικόνα 9 & 10



Εικόνα 10

Αυτό είναι εφικτό μέσω των εξής σταδίων: χρονική ευθυγράμμιση, ευθυγράμμιση επίπεδου, μετασχηματισμός χρόνου - συχνότητας, παραμόρφωση συχνότητας και μετασχηματισμός ακουστικότητας - φιλτράρισμα.<sup>8</sup> Το γνωστικό μοντέλο από την άλλη κάνει πολύπλοκους μη γραμμικούς υπολογισμούς για να υπολογίσει την υποβάθμιση της ποιότητας ομιλίας μέσω της σύγκρισης των ψυχοσωματικών αναπαραστάσεων του σήματος αναφοράς και του υποβαθμισμένου σήματος.



Σχήμα 4

<sup>8</sup> Βλέπε Σχήμα 4

Τα βασικά πρότυπα αξιολόγησης είναι<sup>9</sup>:

- ITU-T P.862: Πρόκειται για το αρχικό πρότυπο με σκοπό την αντικειμενική αξιολόγηση ομιλίας από άκρη σε άκρη για τα δίκτυα στενής ζώνης και κωδικοποιητές ομιλίας.
- ITU-T P.862.1: Βελτίωσε σε σχέση με την προηγούμενη έκδοση τη λειτουργία χαρτογράφησης για την αντιστοίχιση του αποτελέσματος του P.862 σε βαθμολογία MOS.
- ITU-T P.862.2: Ευρείας κλίμακας επέκταση για την εκτίμηση των ευρυζωνικών δικτύων τηλεφωνίας και κωδικοποιητών τηλεφωνίας.
- ITU-T P.862.3: Πρόσθεσε επιπλέον οδηγό εφαρμογής για αντικειμενική αξιολόγηση της ποιότητας φωνής συνοψίζοντας τα πρότυπα P.862, P.862.1 & P.862.2.
- ITU-T P.563 : Αφορά μέθοδο αντικειμενικής αξιολόγησης της ποιότητας φωνής σε εφαρμογές που αφορούν στενά δίκτυα τηλεφωνίας με τη μέθοδο βαθμολόγησης του τελικού και μόνο παραγόμενου σήματος (Δεν πρόκειται για αλγόριθμο αναφοράς)

ITU-T P.862	• Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of <b>narrow-band</b> telephone networks and speech codecs
ITU-T P.862.1	• <b>Mapping function</b> for transforming P.862 raw result scores to MOS-LQO
ITU-T P.862.2	• <b>Wideband</b> extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs
ITU-T P.862.3	• <b>Application guide</b> for objective quality measurement based on Recommendations P.862, P.862.1 and P.862.2
ITU-T P.563	• Single-ended method for objective speech quality assessment in <b>narrow-band</b> telephony applications (no-reference algorithm)

Εικόνα 11

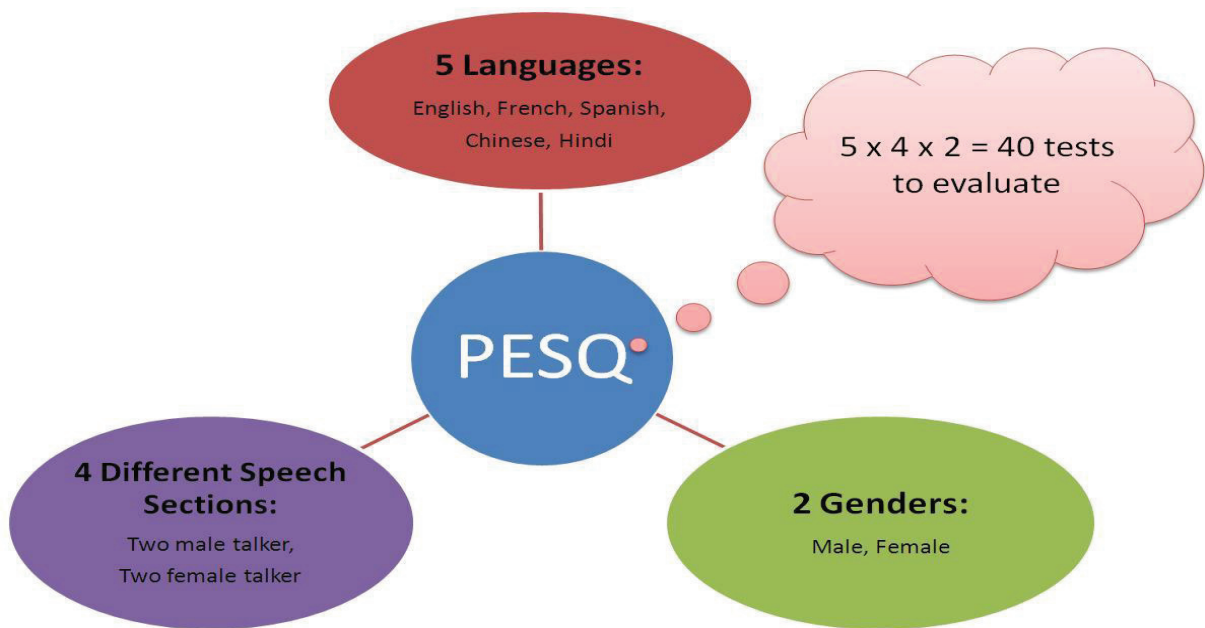
<sup>9</sup> Βλέπε εικόνα 11



Σύμφωνα με την ITU και συγκεκριμένα τη σύσταση P.862 η δοκιμή της PESQ πρέπει να περιλαμβάνει τα ακόλουθα περιεχόμενα<sup>10</sup>:

- Πέντε (5) διαφορετικές Γλώσσες
- Δυο (2) διαφορετικά φύλα (Άνδρας -Γυναίκα)
- Τέσσερα (4) διαφορετικά τμήματα διαλόγου

Συνολικά δηλαδή 40 διαφορετικές μετρήσεις πρέπει να γίνουν για να οριστικοποιηθεί το τελικό αποτέλεσμα.



Σχήμα 5

#### 4.1.4 Επίδοση του αλγόριθμου POLQA

Η μέθοδος POLQA (Perceptual Objective Listening Quality Analysis) αποτελεί την τελευταία τεχνολογία<sup>11</sup> των παρεμβατικών αντικειμενικών μεθόδων αξιολόγησης ποιότητας της φωνής. Αναπτύχθηκε από μια ομάδα ειδικών του ITU-T Study Group 12 την περίοδο 2006 - 2011 οι οποίοι είχαν συμμετάσχει και στην ανάπτυξη των προτύπων P.861 (PSQM), P.862 (PESQ), P.563 (3SQM) και πολλών άλλων ακόμη. Η POLQA μέθοδος οριστικοποιήθηκε το 2011 με

<sup>10</sup> Βλέπε σχήμα 5

<sup>11</sup> Βλέπε σχήμα 6

την δημοσίευση της ITU-T P.863 σύστασης, είναι κατάλληλη για τον έλεγχο ποιότητας δικτύων της σταθερής, κινητής αλλά και μέσω πρωτοκόλλου IP τηλεφωνίας και μπορεί να έχει εφαρμογή σε ευρυζωνικά δίκτυα με ποιότητας φωνής υψηλής ευκρινείας (HD Voice), σε δίκτυα 3G, σε 4G/LTE δίκτυα και φυσικά σε όλα τα συμβατικά και 2G δίκτυα τηλεφωνίας που υποστηρίζουν και οι προγενέστεροι αλγόριθμοι. Επίσης υποστηρίζει τις πιο πρόσφατες κωδικοποιήσεις φωνής και τεχνολογίες μετάδοσης VoIP πράγμα που τον καθιστά ιδανικό για την αξιολόγηση, την βελτιστοποίηση και την παρακολούθηση της ποιότητας φωνής των δικτύων επόμενης γενιάς, έτσι αναμένεται ότι η POLQA/P.863 θα αντικαταστήσει στο εγγύς μέλλον τον ευρέως αυτή τη στιγμή αλγόριθμο PESQ/P.862.

Πιο αναλυτικά οι αναβαθμίσεις του POLQA που διορθώνει πολλές από τις ελλείψεις του PESQ είναι :

- Μπορεί να κάνει μετρήσεις τηλεφωνίας σε δίκτυα ευρείας ζώνης (Wideband Networks)
- Δυνατότητα αξιόπιστων πλέον αποτελεσμάτων χρησιμοποιώντας τον κώδικα EVRC και γενικότερα σε δίκτυα CDMA
- Έχει επιπλέον ευαισθησία σε συγκεκριμένα δίκτυα GSM/WCDMA κάτω από συγκεκριμένες συνθήκες
- Δυνατότητα επιλογής περιορισμών στο VoIP π.χ. μεταβαλλόμενη καθυστέρηση έως 1 second
- Ισχυρές γραμμικές στρεβλώσεις και μεταβολές απολαβής π.χ. το σχήμα συχνότητας της τηλεφωνικής σύνδεσης
- Περιορισμοί της υποβάθμισης ποιότητας μέσω συσκευών επεξεργασίας για την βελτίωση της φωνής π.χ. θόρυβος γραμμής ή ο αυτόματος έλεγχος απολαβής

- Δυνατότητα λειτουργίας και στα τρία διαθέσιμα εύρη ζώνης Narrowband(300-3400Hz, Wideband(100-7000Hz και Super Wideband(50-14000Hz)
- Λειτουργεί κάτω από διαφορετικά περιβάλλοντα θορύβου όπως π.χ. ένας δρόμος, ένα εστιατόριο ή ο θόρυβος μια μηχανής
- Ικανότητα λειτουργίας σε όλους τους νέους εμπορικούς και τυποποιημένους κωδικοποιητές που χρησιμοποιούνται σε διαφορετικές τεχνολογίες (GSM/WCDMA, LTE, VoIP) όπως π.χ. οι AMR-WB, EVRC, AAC, Skype/SLIK, G.711, G.729 κ.λ.π.
- Επιτρέπει το συνδυασμό μετρήσεων διαφόρων διεπαφών όπως π.χ. από ηλεκτρική σε ηλεκτρική, από ακουστική σε ακουστική, από ηλεκτρική σε ακουστική και από ακουστική σε ηλεκτρική
- Παροχή δυνατότητας μέτρησης τερματικών συσκευών άλλα και hands-free εφαρμογών
- Δυνατότητα μετρήσεων σε πολλές περισσότερες γλώσσες ομιλίας οι οποίες είναι Αγγλική (Αμερικάνικη και Βρετανική), Κινέζικη, Γερμανική, Ιταλική, Γαλλική, Τσέχικη, Σουηδική, Γιαπωνέζικη, Βαυαρική και ελβετική γλώσσα.

Αν και πρόκειται για αλγόριθμο τελευταίας γενιάς χωρίς ακόμα να έχει μπει πλήρως στην αγορά λόγω ανακολουθίας αναβάθμισης των συστημάτων τηλεφωνίας από τους παρόχους η POLQA έχει κάποιους πολύ αυστηρούς περιορισμούς όσον αφορά την χρήση της. Οπότε δεν μπορεί να χρησιμοποιηθεί η POLQA όταν υπάρχουν οι παρακάτω λόγοι:

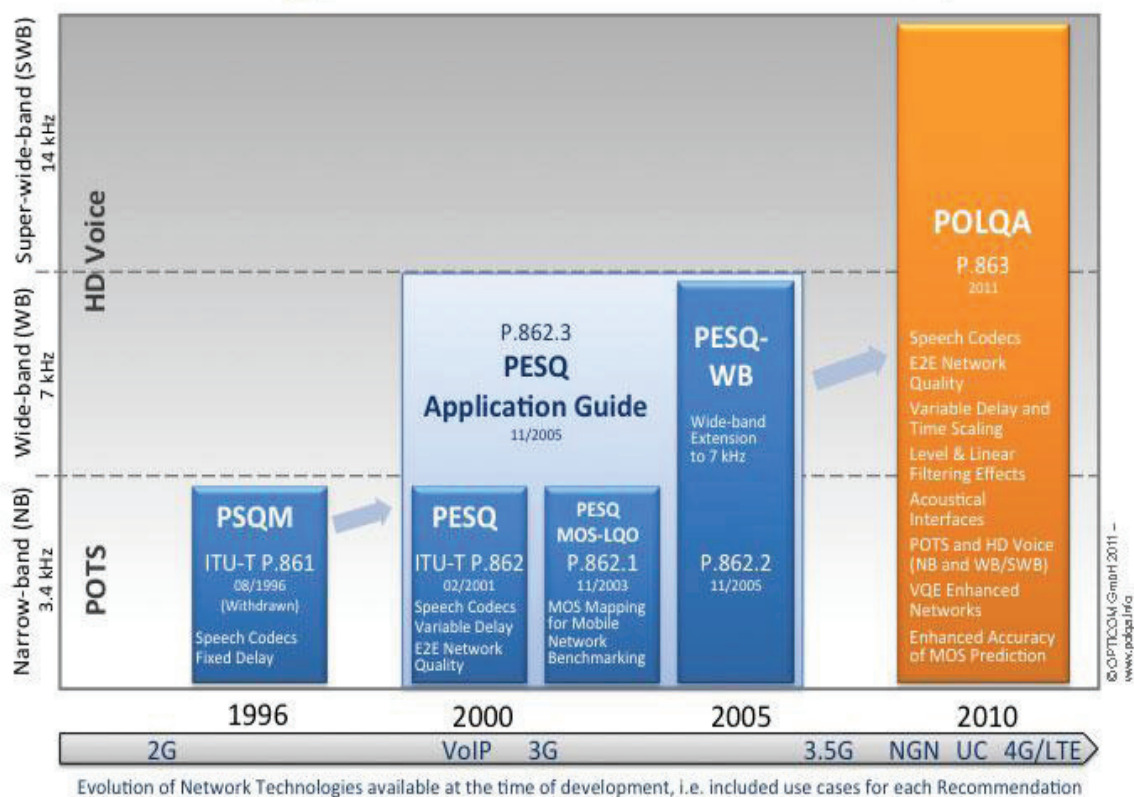
- ♦ Διαφορετικές διαστάσεις κλίμακας ποιότητας ομιλίας
- ♦ Μέτρηση ποιότητας ομιλίας σε μια ολόκληρη τηλεφωνική συνδιάλεξη (Η POLQA έχει σαν σκοπό να μετράει ομιλίες διάρκειας από 6 έως 12 δευτερόλεπτα)
- ♦ Μέτρηση μουσικής συμπεριλαμβανομένων των πολυμέσων

- ♦ Μέτρηση της απόδοσης των συσκευών βελτίωσης φωνής (Όπως οι καταστολείς θορύβου)
- ♦ Μετρήσεις άλλων τεχνολογιών ή παρελκόμενων όπως συσκευές καταγραφής φωνής ή εφαρμογές όχι σχετικές με την τηλεφωνία όπως επαγγελματικές κινητές ραδιοζεύξεις.

Μπορεί επίσης να χρησιμοποιηθεί αλλά με επιφυλάξεις ως προς τα αποτελέσματα καθώς δεν έχουν πιστοποιηθεί οι παράμετροι αυτοί στις εξής περιπτώσεις:

- ♦ Διαφορετικές γλώσσες ομιλίας από τις πιστοποιημένες όπως π.χ. η Αραβική γλώσσα
- ♦ Μετρήσεις μεγαλύτερων χρονικά ομιλιών από το προβλεπόμενο διάστημα
- ♦ Μετρήσεις ήχου πολλών διαστάσεων για την αξιολόγηση της ποιότητας ομιλίας

## Evolution of ITU-T Recommendations for Voice Quality Testing (P.86x - Full Reference MOS-LQO)



Σχήμα 6

## 4.2 Μη Παρεμβατικές αντικειμενικές μέθοδοι

Οι μη παρεμβατικές αντικειμενικές μέθοδοι αξιολόγησης της ποιότητας ομιλίας παρακολουθούν διαρκώς την τελική ποιότητα της ομιλίας που φτάνει στον πελάτη ή την ποιότητα ομιλίας που υπάρχει σε ένα συγκεκριμένο κόμβο του δικτύου. Οι μέθοδοι αυτοί χρησιμοποιούν σήματα στο δίκτυο για να κάνουν προβλέψεις για την ποιότητα ομιλίας. Η επεξεργασία των ενδοϋπηρεσιακών αυτών σημάτων αντι για δοκιμές ερεθισμού και αντίληψης της παρεχόμενης ποιότητας σημαίνει ότι οι μη παρεμβατικές αντικειμενικές μέθοδοι μπορούν να χρησιμοποιηθούν μόνο για να προβλέψουν την παρεχόμενη ποιότητα ομιλίας που λαμβάνει ο τελικός χρήστης. Οι παρεμβατικές μέθοδοι μπορεί να είναι παράμετροι του δικτύου ή παράμετροι που βασίζονται στο σήμα της ομιλίας. Οι παράμετροι του δικτύου δεν χρησιμοποιούν καθόλου το σήμα φωνής που διέρχεται από το δίκτυο αλλά προβλέπουν την παρεχόμενη ποιότητα ομιλίας που λαμβάνει ο χρήστης μόνο από τεχνικές ρυθμίσεις αυτού και ρυθμίσεις μεταφοράς του σήματος. Οι παράμετροι που βασίζονται στο σήμα της ομιλίας χρειάζονται για να προβλέψουν την ποιότητα αυτής ένα αρχικά μεταδιδόμενο σήμα το οποίο να μην είναι καθόλου υποβαθμισμένο ποιοτικά. Σε περιπτώσεις που υπάρχει μεγάλη υποβάθμιση θα είναι δυνατό να επηρεαστεί σημαντικά το αποτέλεσμα της μέτρησης. Γίνεται αντιληπτό λοιπόν ότι οι μη παρεμβατικές μέθοδοι αξιολόγησης έχουν πολύ μικρότερη ακρίβεια απ' ότι οι παρεμβατικές μέθοδοι αξιολόγησης. Παρότι όμως είναι λιγότερο ακριβείς οι μη παρεμβατικές μέθοδοι διαδραματίζουν ένα σημαντικό ρόλο στο στάδιο ανάπτυξης ενός δικτύου και μέχρις ότου αυτό ολοκληρωθεί. Όταν το δίκτυο όμως είναι έτοιμο λειτουργικά και αρχίσει να χρησιμοποιείται από τους χρηστές, οι παρεμβατικές μέθοδοι προτείνονται για την αξιολόγηση της παρεχόμενης ποιότητας φωνής από άκρη σε άκρη του δικτύου όπως και για την επίλυση τυχόν προβλημάτων στην απόδοση του.

### 4.2.1 Επίδοση του αλγόριθμου 3SQM

Η μέθοδος 3SQM ή αλλιώς Single Speech Quality Measure η οποία προδιαγράφεται στην σύσταση της ITU P.563 μπορεί να εφαρμοστεί για να προβλέψει την ποιότητα της ομιλίας χωρίς ξεχωριστό σήμα αναφοράς. Για το λόγο αυτό συνίσταται για μη παρεμβατική αξιολόγηση της ποιότητας ομιλίας, ζωντανή παρακολούθηση του δικτύου και αξιολόγηση χρησιμοποιώντας άγνωστες πηγές ομιλίας στο άλλο άκρο της τηλεφωνικής σύνδεσης.

Η αξιολόγηση της ποιότητας φωνής με τη μέθοδο αυτή απαιτεί πολλά βήματα. Πριν την αξιολόγηση της ποιότητας του σήματος της φωνής το σήμα πρέπει να υποστεί μια προεπεξεργασία. Τα σημαντικά βήματα της προεπεξεργασίας είναι το φιλτράρισμα για την αποκοπή ορισμένων συχνοτήτων, η ρύθμιση της στάθμης της ομιλίας και ο διαχωρισμός σε φωνητικά και μη φωνητικά μέρη μέσω ανίχνευσης δραστηριότητας φωνής (VAD, Voice Activity Detection).

Σε δεύτερη φάση, εξάγονται από το σήμα η παραμόρφωση και κάποιες παράμετροι της ομιλίας. Τρία βασικά λειτουργικά τμήματα ανάλυσης που αντιστοιχούν στις κύριες ομάδες των παραμορφώσεων οι οποίες ορίζονται στην P.563, βοηθούν στην ταξινόμηση των σημάτων και των παραμορφώσεών τους.

Οι βασικές κατηγορίες παραμορφώσεις είναι οι εξής:

Μη φυσική ροή του λόγου, βασική ποιότητα ομιλίας, τεχνητή ή ρομποτική φωνή, αφύσικη φωνή σαν βόμβος, ισχυρός προσθετικός θόρυβος, θόρυβος υποβάθρου, τμηματικά χαμηλό SNR (Signal-to-noise ratio, λόγος σήματος προς θόρυβο), διακοπές και αποκοπές.

Στο τέλος και ύστερα από ανάλυση, καθορίζεται η τάξη παραμόρφωσης και υπολογίζεται η τιμή της γνωμοβαθμολογίας MOS.



**Εικόνα 12**

Ένα από τα σημαντικότερα πλεονεκτήματα του P.563 σε σχέση με άλλες προσεγγίσεις μέτρησης είναι ότι δεν χρειάζεται να γίνουν υποθέσεις σχετικά με το υπό δοκιμή δίκτυο ή για το είδος παραμόρφωσης που μπορεί να συμβεί σε ένα συγκεκριμένο σενάριο δοκιμής τηλεφωνίας. Η μόνη προϋπόθεση που απαιτεί το μοντέλο είναι η επιστημονική γνώση σχετικά με το πώς παράγεται η ανθρώπινη ομιλία και το πώς αυτή γίνεται αντιληπτή από τον άνθρωπο. Ο αλγόριθμος P.563 περιλαμβάνει στον υπολογισμό τις επιπτώσεις υποβάθμισης

που σχετίζονται τόσο με τη μεταγωγή πακέτων όσο και με το ίδιο το σήμα (θόρυβος και παραμορφώσεις που προκαλούνται από τις διαδικασίες κωδικοποίησης).

Το μοντέλο παραμόρφωσης που βασίζεται σε αυτή τη γνώση είναι τόσο γενικό όσο επιτρέπεται και ως εκ τούτου μπορεί να είναι ανεξάρτητο από τις τρέχουσες και μελλοντικές εφαρμογές της τηλεφωνίας και τα δίκτυα.

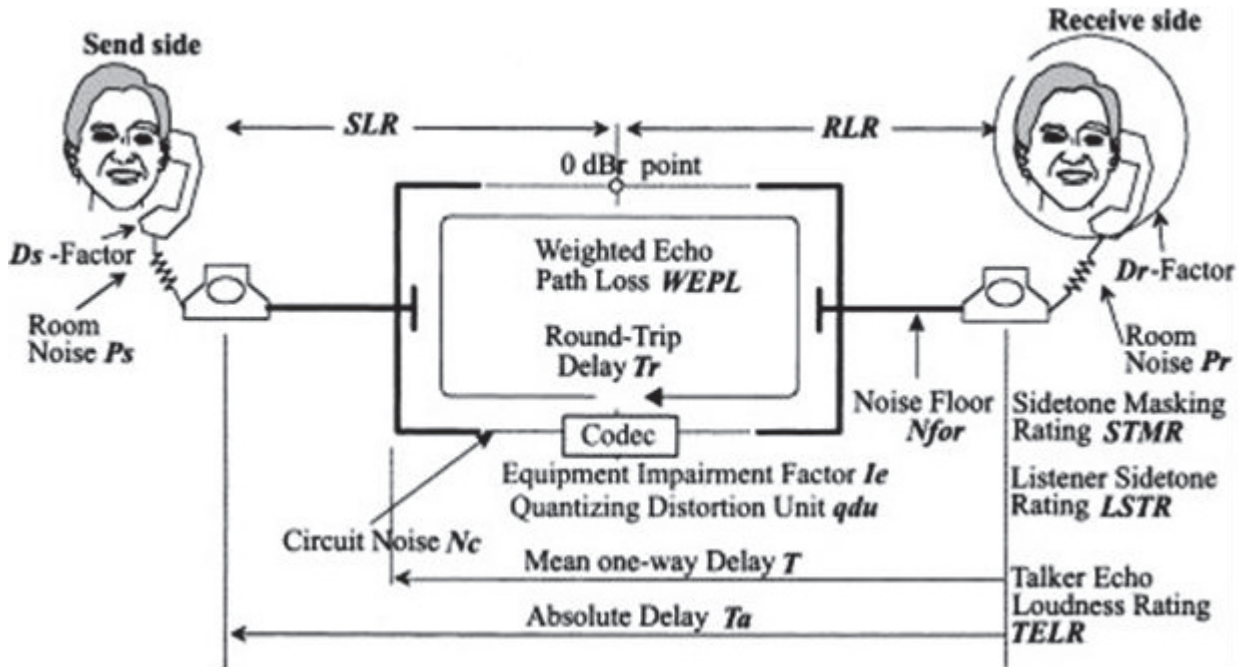
#### **4.2.2 Επίδοση του αλγόριθμου E-model**

Το E-model αναφέρεται στη συσταση ITU-T G.107, είναι ένα εργαλείο σχεδιασμού μετάδοσης που παρέχει πρόβλεψη της αναμενόμενης ποιότητας φωνής, όπως αυτή γίνεται αντιληπτή από έναν τυπικό χρήστη τηλεφώνου, για μια ολοκληρωμένη τηλεφωνική σύνδεση υπό συνθήκες ομιλίας (δηλαδή συνομιλίας μετάδοσης ήχου από το στόμα στο αυτί). Αυτός ο αλγόριθμος λαμβάνει υπ' όψιν ένα ευρύ φάσμα τηλεφωνίας, ιδίως την κακή μετάδοση λόγω του χαμηλού ρυθμού διαμεταγωγής κωδικοποίησης σε συσκευές και την μονόδρομη καθυστέρηση, καθώς και τα «κλασικά» προβλήματα τηλεφωνίας όπως είναι ο θόρυβος και η ηχώ. Μπορεί να εφαρμοστεί για την εκτίμηση της ποιότητας ομιλίας των ενσύρματων και ασύρματων δικτύων, με βάση την μεταγωγή του κυκλώματος και την μεταγωγή των πακέτων της τεχνολογίας.

Αυτό το υπολογιστικό μοντέλο βασίζεται στην μοντελοποίηση των αποτελεσμάτων από ένα μεγάλο αριθμό υποκειμενικών δοκιμών που έγιναν στο παρελθόν. Είναι χρήσιμο στους μηχανικούς που σχεδιάζουν μια μετάδοση, και βοηθά ώστε να διασφαλιστεί ότι οι χρήστες θα είναι ικανοποιημένοι με την απόδοσή της από άκρο σε άκρο. Η κύρια έξοδος του μοντέλου είναι μια κλιμακωτή αξιολόγηση της ποιότητας μετάδοσης. Ένα ακόμη σημαντικό



χαρακτηριστικό του μοντέλου αυτού είναι η χρήση παραγόντων υποβάθμισης της μετάδοσης, που αντανakλούν τις επιδράσεις των σύγχρονων συσκευών επεξεργασίας σήματος.



Σχήμα 7: E-model

Το E-model βασίζεται στην μοντελοποίηση των αποτελεσμάτων από ένα μεγάλο αριθμό υποκειμενικών δοκιμών που έγιναν στο παρελθόν σε ένα ευρύ φάσμα των παραμέτρων μετάδοσης. Η απόδοση του υπολογισμού αυτού του αλγόριθμου γίνεται μέσω μιας βαθμωτής τιμής της βαθμολογίας η οποία είναι γνωστή ως «Παράγοντας Αξιολόγησης Μεταφοράς R». Το R μπορεί να μετατραπεί σε άλλα μέτρα, όπως η ποιότητα της τιμής της MOS και μετράται σαν Ποσοστό καλό ή καλύτερο (GOB) ή Ποσοστό Κακό ή χειρότερα (PoW). Ωστόσο, πρέπει να δίνεται προσοχή όταν συγκρίνουμε αυτά τα μέτρα γιατί μετατρέπονται με τις αξίες της MOS, σε ποσοστά επί των τιμών GOB ή PoW από άλλες πηγές, οι οποίες μπορεί να μην έχουν ληφθεί υπό παρόμοιες συνθήκες.

Η σχέση μεταξύ των διαφόρων παραγόντων δυσλειτουργίας και R δίνεται από την εξίσωση:

$$R = R_0 - I_s - I_d - I_{e,eff} + A$$

Όπου,

Ο όρος  $R_0$  εκφράζει τη βασικό λόγο σήματος-προς-θόρυβο (επίπεδο λαμβανόμενης ομιλίας σε σχέση με το θόρυβο του κυκλώματος και ακουστική),

Ο όρος  $I_s$  αντιπροσωπεύει όλες τις βλάβες που συμβαίνουν σχεδόν ταυτόχρονα με το σήμα της φωνής, όπως: το πολύ δυνατό επίπεδο του λόγου, η μη-βέλτιστη τιμή πλάγιου τόνου (STMR), ο θόρυβος κβαντοποίησης (qdu), κλπ.

Ο όρος  $I_d$  συνοψίζει όλα τα προβλήματα λόγω της καθυστέρησης και του εφέ της ηχώ.

Ο όρος  $I_{e,eff}$  είναι ένας «αποτελεσματικός παράγοντας απομείωσης εξοπλισμού», ο οποίος αντιπροσωπεύει βλάβες που προκαλούνται από τους κωδικοποιητές με χαμηλό ρυθμό διαμεταγωγής. Περιλαμβάνει επίσης απομείωση λόγω πακέτων απώλειας τυχαίας κατανομής.

Οι τιμές του  $I_e$  για συγκεκριμένους κωδικοποιητές χωρίς απώλεια πακέτων δίνεται στην ITU-T Rec. G.113 App. I, και αυτές οι αξίες μετατρέπονται σε  $I_{e,eff}$  σε περίπτωση τυχαίας απώλειας πακέτων, χρησιμοποιώντας το μοντέλο E-model.

Ο όρος  $A$  είναι ένας «παράγοντας πλεονεκτήματος», ο οποίος δίνει δυνατότητα πρόσβασης για ορισμένα συστήματα σε σχέση με τα συμβατικά συστήματα, εμπορίας ποιότητας φωνής για ευκολία. Ενώ όλοι οι άλλοι παράγοντες απομείωσης αφαιρούνται από το βασικό λόγο σήματος-προς-θόρυβο  $R_0$ , το  $A$  προστίθεται και έτσι αντισταθμίζει άλλες βλάβες σε ένα ορισμένο ποσό. Μπορεί να χρησιμοποιηθεί για να ληφθεί υπ' όψιν το γεγονός ότι ο χρήστης θα ανεχτεί κάποια μείωση στην ποιότητα μετάδοσης σε αντάλλαγμα με το «πλεονέκτημα της πρόσβασης». Παραδείγματα τέτοιων πλεονεκτημάτων είναι τα ασύρματα και κινητά συστήματα ή συνδέσεις που είναι δύσκολο να φτάσουν σε περιοχές μέσω δορυφόρου.

Οι τιμές του R κυμαίνονται μεταξύ από 0 ≤ R < 100, με τις υψηλότερες τιμές να δείχνουν υψηλότερη ποιότητα ομιλίας. Στον ακόλουθο πίνακα φαίνονται οι βαθμολογίες του E-model σε βαθμολογίες R σε κατηγορίες ποιότητας μετάδοσης ομιλίας και για την ικανοποίηση των χρηστών.

<b>Range of E-model Rating R</b>	<b>Speech transmission quality category</b>	<b>User satisfaction</b>
90 ≤ R < 100	Best	Very satisfied
80 ≤ R < 90	High	Satisfied
70 ≤ R < 80	Medium	Some users dissatisfied
60 ≤ R < 70	Low	Many users dissatisfied
50 ≤ R < 60	Poor	Nearly all users dissatisfied
NOTE 1 □ Connections with E-model Ratings R below 50 are not recommended. NOTE 2 □ Although the trend in transmission planning is to use E-model Ratings R, equations to convert E-model Ratings R into other metrics, e.g. %MOS, %GoB, PoW can be found in ITU-T Rec. G.107 Annex B		

**Πίνακας 6: Κατηγορίες της ποιότητας μετάδοσης ομιλίας**

Το E-model δημοσιεύτηκε για πρώτη φορά το 1998 και έκτοτε η ITU-T το βελτιώνει συνεχώς αναθεωρώντας το και εμπλουτίζοντάς το κάθε 1-2 χρόνια. Το 2000, συμπεριλήφθηκαν στους υπολογισμούς του μοντέλου οι επιπτώσεις του θορύβου του χώρου στην πλευρά της εκπομπής και η παραμόρφωση κβαντοποίησης. Στην έκδοση του 2002 συμπεριλήφθηκε με έναν παραμετρικό τρόπο, η υποβάθμιση που οφείλεται στην τυχαία απώλεια πακέτων για διαφορετικούς κωδικοποιητές. Το 2003 βελτιώθηκε η μοντελοποίηση της ποιότητας στην περίπτωση χαμηλού επιπέδου πλάγιου τόνου της συσκευής του ομιλητή.

Η έκδοση του 2005 επέτρεψε πιο ακριβείς προβλέψεις για την ποιότητα των codecs κάτω από συνθήκες απώλειας πακέτων.

Στην τελευταία έκδοση της σύστασης περιλαμβάνεται ένα παράρτημα όπου περιγράφεται ένα προσωρινό πλαίσιο παραγόντων υποβάθμισης για ευρυζωνικά συστήματα μετάδοσης ομιλίας και μια εφαρμογή αναφοράς του μοντέλου.

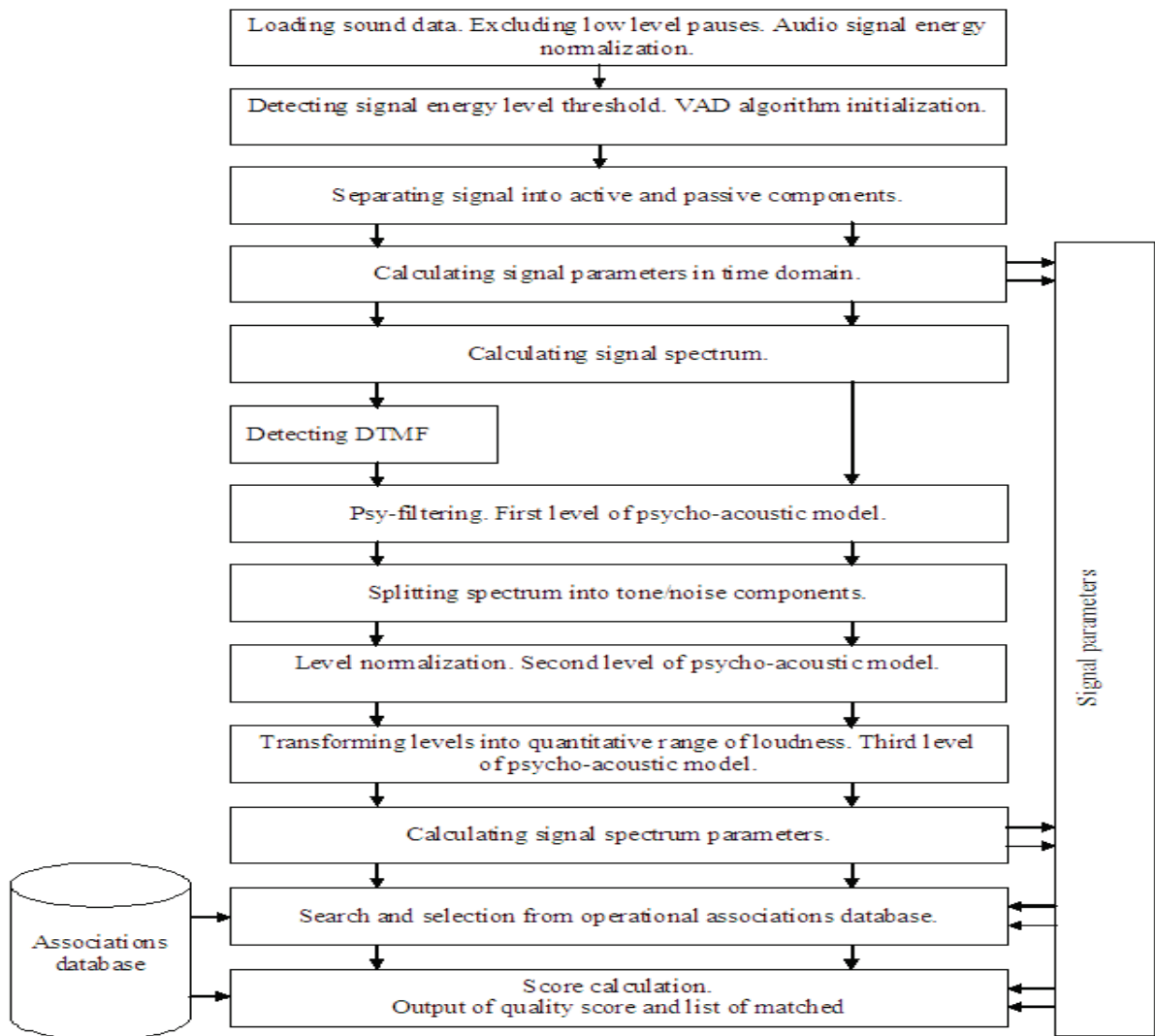
### **4.2.3 Επίδοση του αλγόριθμου NiQA**

Ο NiQA (Non-Intrusive voice Quality Analyzer) είναι ένας αλγόριθμος για τη μέτρηση της ποιότητας της φωνής. Μπορεί ακόμη να κάνει δοκιμαστικές κλήσεις χωρίς φωνή αναφοράς σε μακρινές τηλεφωνικές συνδέσεις. Αυτό είναι θετικό για τις διεθνείς κλήσεις, όπου οι φωνές αναφοράς είναι δύσκολο να επιτευχθούν. Το σύστημα κάνει ακριβείς προβλέψεις σε ένα ευρύ φάσμα των στρεβλώσεων του δικτύου.

Για την παρακολούθηση της ποιότητας της ομιλίας κάτω από πραγματικές συνθήκες απαιτείται μόνο ένα σημείο μέτρησης. Η βάση δεδομένων που ρυθμίζει τα αποτελέσματα της βαθμολογίας του NIQA περιλαμβάνει ένα ευρύ φάσμα των συνθηκών του δικτύου και των προβλήματα όπως οι κωδικοποιητές ομιλίας χαμηλού ρυθμού διαμεταγωγής, οι παύσεις ομιλίας, ο θόρυβος του περιβάλλοντος κ.λπ.

Η προσέγγιση του αλγόριθμου NIQA βασίζεται σε μια βάση δεδομένων η οποία αποτελείται από προκαθορισμένα βήματα που ονομάζονται συσχετισμοί. Κάθε συσχετισμός ανταποκρίνεται σε μια ομάδα αρχείων που εμπεριέχει ακριβείς μετρήσεις της ποιότητας του ήχου και των κοινών παραμέτρων που την υποβαθμίζουν. Για κάθε συσχετισμό ο NIQA υπολογίζει και αποθηκεύει μια κατανομή των τιμών των παραμέτρων.

Βασικοί αλγόριθμοι που μας δείχνουν πως ο NIQA μπορεί να δώσει αποτελέσματα ποιότητας ήχου παρουσιάζονται στο σχήμα 8.



Σχήμα 8

Κατά τη διαδικασία που φορτώνεται το σήμα του ήχου από το σύστημα αποκλείονται όλα τα αποσπάσματα με χαμηλό ενεργειακό επίπεδο (σύμφωνα με το όριο). Τα αποσπάσματα που απορρίπτονται αντιστοιχούν σε «απόλυτη σιωπή» και θεωρούνται άνευ σημασίας για τη λήψη ήχου βαθμολογία ποιότητας.

Κατά την επόμενη φάση το σήμα διαιρείται σε πλαίσια που χρησιμοποιούνται σε αλγόριθμο ανίχνευσης φωνητικής δραστηριότητας (VAD). Το σύστημα υπολογίζει τις τιμές ενέργειας για κάθε πλαίσιο που αυξάνει την ακρίβεια των VAD. Με τη βοήθεια του VAD ο αλγόριθμος

χωρίζει το σήμα σε ενεργά και ανενεργά συστατικά που υποβάλλονται σε επεξεργασία χωριστά.

Με διακριτού μετασχηματισμού συνημίτονο (DCT) το σύστημα αποκτά σήμα φάσματος και ελέγχει τα ενεργά εξαρτήματα για την παρουσία DTMF και στη συνέχεια αποκλείει τα πλαίσια που είναι παρόμοια με DTMF από την περαιτέρω επεξεργασία.

Στο επόμενο στάδιο εφαρμόζει το πρώτο επίπεδο του ψυχο-ακουστικού μοντέλου με το φάσμα του σήματος. Αυτό το μοντέλο ελέγχει διαφορετικούς τύπους συγκάλυψης (συμπεριλαμβανομένης της προ-επικάλυψης και μετά συγκάλυψης). Σύμφωνα με σαφείς αιχμές της ενέργειας του φάσματος, το σύστημα χωρίζει το σήμα στον τόνο και τα συστατικά του θορύβου.

Στο δεύτερο επίπεδο του ψυχο-ακουστικού μοντέλου εκτελεί ομαλοποίηση της ενέργειας του σήματος. Τα επίπεδα της ενέργειας μετατρέπονται σε επίπεδα έντασης στο 1kHz. Στο τρίτο επίπεδο του ψυχο-ακουστικού μοντέλου μετατρέπει τα επίπεδα της έντασης σε διάφορες ποιότητες ανιχνεύσιμης έντασης που επιτρέπουν να αγνοήσουμε τις αλλαγές του ήχου του σήματος, οι οποίες δεν αναγνωρίζονται από το ανθρώπινο αυτί.

Το επόμενο βήμα είναι να χωριστεί το φάσμα του σήματος σε ζώνες που είναι κρίσιμες για την ανθρώπινη αντίληψη από το αυτί και να υπολογίσει τις παραμέτρους τόσο εντός όσο και έξω από τις ζώνες. Βάσει των υπολογιζόμενων παραμέτρων του σήματος το σύστημα επιλέγει τις πιο όμοιες ενώσεις από τη βάση δεδομένων και τις εκτελεί ώστε να ταιριάζουν. Σύμφωνα με τις επιλεγμένες ενώσεις το σύστημα καθορίζει πόσο η κάθε μια από αυτές επηρεάζει την συνολική ποιότητα και στη συνέχεια δημιουργεί την τελική τιμή της ποιότητας φωνής ως συνδυασμό των τιμών για επιλεγμένες απο τις ενώσεις και σύμφωνα με το ειδικό βάρος της καθεμιά από αυτές.

## ΚΕΦΑΛΑΙΟ 5

### ΤΕΛΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ - ΣΥΓΚΡΙΣΗ

Όπως γίνεται εύκολα αντιληπτό από τα προηγούμενα κεφάλαια, υπάρχουν πάρα πολλές παράμετροι οι οποίες είναι ικανές να επηρεάσουν την ποιότητα ήχου της ομιλίας που αντιλαμβάνεται ο τελικός χρήστης της τηλεφωνίας. Βασικό ρόλο παίζουν σε αυτό πολλές τεχνικές παράμετροι όπως οι κωδικοποιητές φωνής ή τα μέσα μετάδοσης της ομιλίας όπως π.χ. η τηλεφωνική συνομιλία μέσω ίντερνετ που εξαρτάται εν μέρει από το ρυθμό διαμεταγωγής των δεδομένων ή μια απλή τηλεφωνική σύνδεση η οποία εξαρτάται από τον πλέον απλό παράγοντα πιθανής υποβάθμισης της ποιότητας του ήχου, αυτό τη τηλεφωνικής συσκευής και της ποιότητας κατασκευής που αυτή έχει.

Όσον αφορά το θέμα της παρούσας εργασίας όμως αυτό που πλέον έχει άμεσο αντίκτυπο σε όλα τα προηγούμενα είναι η σωστή αποτίμηση της ενδεχόμενης υποβάθμισης ποιότητας του αρχικού παραγόμενου ήχου ομιλίας διότι σε αντίθετη περίπτωση τα αποτελέσματα είναι πιθανόν να μας οδηγήσουν σε εντελώς λάθος συμπεράσματα με άμεση συνέπεια την ολική υποβάθμιση και απαξίωση του τηλεφωνικού δικτύου ή μέρους αυτού.

Σωστή μέθοδος που εμπεριέχει όλα τα σφάλματα ή ελλείψεις των προηγούμενων σαφώς δεν υπάρχει και θα ήταν εντελώς λάθος αν οποιοσδήποτε προσπαθούσε να υπονοήσει κάτι τέτοιο. Είναι επίσης σαφές ότι από την στιγμή που αναφερόμαστε στην αξιολόγηση της ποιότητας ομιλίας, όχι βάση των τεχνικών προδιαγραφών που το υπό δοκιμή ηχητικό δείγμα λαμβάνει μέρος στο εκάστοτε πείραμα, αλλά βάση της ποιότητας που αντιλαμβάνεται ο χρήστης μιλάμε για κάτι υποκειμενικό σε τέτοια όμως έκταση δειγματοληψίας που να θεωρείται ευρέως αποδεκτό άρα και εντέλει αντικειμενικό.

Ως εκ τούτου λαμβάνοντας υπ' όψιν τα παραπάνω θα μπορούσαμε να θεωρήσουμε ως την πλέον αξιόπιστη μέθοδο αξιολόγησης αυτή της υποκειμενικής αξιολόγησης της ποιότητας ομιλίας. Άλλωστε κάτι τέτοιο το επιβεβαιώνει η ίδια η λογική λειτουργίας της αντικειμενικής αξιολόγησης στην οποία παρότι δεν συμμετέχει ενεργά σε κανένα σημείο της διαδικασίας αυτής ο ανθρώπινος παράγοντας, η κατεύθυνση της για την εξαγωγή αποτελεσμάτων είναι η όσο το δυνατόν καλύτερη προσομοίωση της υποκειμενικής μεθόδου αξιολόγησης της ποιότητας ομιλίας. Αυτό λοιπόν μας αποδεικνύει ότι η μόνη σωστή μέθοδος αξιολόγησης είναι αυτή η οποία ταιριάζει απόλυτα ή όσο το δυνατόν περισσότερο με δεδομένα που το πείραμα μας περιλαμβάνει όπως και τα κριτήρια που αυτό θέτει.

Αναλύοντας λοιπόν τις μεθόδους μια προς μια έχουμε τα εξής:

Οι δοκιμές αξιολόγησης συνομιλίας περιλαμβάνονται στην υποκειμενική μέθοδο αλλά έχουν σαν όρια λειτουργίας την αξιολόγηση σε πραγματικό χρόνο της ποιότητας της ομιλίας που λαμβάνεται γνωρίζοντας της ίδια στιγμή την ποιότητα της ομιλίας που αρχικά παράχθηκε από το ίδιο δείγμα ήχου. Αυτό έχει σαν αποτέλεσμα η αξιολόγηση να βαθμολογεί το σύστημα σε πραγματικό χρόνο μεν αλλά με καθόλου σταθερά σημεία αναφοράς ως προς της σύγκριση των αποτελεσμάτων αυτών με άλλα αποτελέσματα από άλλες πειραματικές μεθόδους. Επίσης η μέθοδος αυτή έχει αρκετά μεγάλο κόστος όπως επίσης είναι και αρκετά χρονοβόρα. Έτσι η δοκιμή αξιολόγησης της συνομιλίας λοιπόν προτείνεται σε περιπτώσεις μόνο που ο διεξάγων το πείραμα ενδιαφέρεται για μια στιγμιαία αξιολόγηση του τηλεφωνικού συστήματος ή μέρους αυτού με το καλύτερο δυνατό παραγόμενο αποτέλεσμα χωρίς να τον απασχολεί το κόστος απόκτησης αυτού ή του χρόνου απόκτησης του. Επειδή το αποτέλεσμα αυτό όμως αν και θεωρητικά βάση της MOS πρόκειται για ότι πιο κοντινό στην πραγματικότητα που θέλουμε να προσεγγίσουμε, η αδυναμία σύγκρισης των αποτελεσμάτων αυτών με



οποιοδήποτε άλλο σύστημα και φυσικά το αρκετά μεγάλο κόστος που απαιτεί, το καθιστούν μη εμπορικά εκμεταλλεύσιμο.

Εν αντιθέσει με τις δοκιμές αξιολόγησης συνομιλίας, οι δοκιμές αξιολόγησης της ακρόασης έχουν ένα συγκριτικό πλεονέκτημα, αυτό της δυνατότητας κοινών σημείων αναφοράς με άλλες μελέτες και δυνατότητα πλέον εξαγωγής συγκριτικών αποτελεσμάτων. Το σημείο αναφοράς είναι το αρχικό δείγμα ήχου ή ομιλίας το οποίο είναι ηχογραφημένο κάτω από συγκεκριμένες συνθήκες που το πείραμα προϋποθέτει και οι συμμετέχοντες στο πείραμα εξάγουν τελικά συμπεράσματα αξιολογώντας το τελικό δείγμα βάση της κλίμακας αξιολόγησης MOS. Με αυτό τον τρόπο το αρχικό δείγμα δεν μπορεί να επηρεάσει με κανέναν τρόπο τους συμμετέχοντες αφού σε κάθε περίπτωση πρόκειται για το ίδιο δείγμα σε όλα τα πειράματα. Οι δοκιμές αξιολόγησης έχουν το συγκριτικό πλεονέκτημα της απόλυτης βαθμολογίας σε ότι αφορά την αξιολόγηση της ποιότητας ομιλίας και είναι το καλύτερο δυνατό πείραμα για την εξαγωγή συμπερασμάτων σε ότι αφορά αυτή. Έχει όμως και το μειονέκτημα του πολύ υψηλού κόστους που απαιτεί για να πραγματοποιηθεί αυτή η έρευνα, (αρκετά άτομα που συμμετέχουν, συγκεκριμένες εγκαταστάσεις βάση προτύπου, αρκετά μεγάλος χρόνος πραγματοποίησης του πειράματος κ.λπ.) ενός κόστους που δεν έχουν όλοι οι πάροχοι τηλεφωνικών δικτύων ή παρεμφερούς αντικειμένου δυνατότητα να ανταποκριθούν για να μελετήσουν την παρεχόμενη ποιότητα ήχου και εν τέλει εμπειρίας που παρέχουν στους πελάτες τους.

Αυτό εν μέρει το λύνει η δοκιμή δημοσκοπικής έρευνας όπου χωρίς ουσιαστικά κάποιο σοβαρό κόστος οι ίδιες εταιρίες μετρούν σε μεταφορική έννοια την ποιότητα εμπειρίας που λαμβάνουν οι πελάτες τους. Μεταφορικά διότι ουσιαστικά δεν πρόκειται για κάποια μέτρηση σε πείραμα βάση προδιαγραφών οποιουδήποτε προτύπου αλλά για τηλεφωνικά ερωτηματολόγια που ζητούν από τους χρήστες του δικτύου να αξιολογήσουν όπως εκείνοι

νομίζουν την ποιότητα του ήχου που λαμβάνουν. Αυτή η μέτρηση σε καμιά περίπτωση δεν μπορεί να αξιολογηθεί ουσιαστικά για την βελτίωση ή αναβάθμιση του δικτύου αφού δεν έχει καμία τεχνική ή τεχνολογική προδιαγραφή αλλά πρόκειται για μια στιγμιαία αξιολόγηση του παρεχόμενου τηλεφωνικού δικτύου. Από τα προηγούμενα λοιπόν προκύπτει ότι οι δοκιμές αυτές μπορούν να χρησιμοποιηθούν μόνο σε υφιστάμενα δίκτυα που είναι σε λειτουργία και όχι σε δίκτυα που είναι σε φάση κατασκευής, ανάπτυξης ή δοκιμών πριν την έναρξη λειτουργίας.

Το πρόβλημα σε όλες τις παραπάνω λύσεις λοιπόν ήρθαν να δώσουν οι αντικειμενικές αλγοριθμικές μέθοδοι αξιολόγησης τις οποίες ανέπτυξαν διάφορες εταιρίες που συνεργάστηκαν για το σκοπό αυτό. Η πρώτη βασική μέθοδος που χρησιμοποιήθηκε ευρέως είναι η PSQM ή αλλιώς μέτρηση ποιότητας ήχου που αντιλαμβανόμαστε, η οποία και προσπάθησε να προσομοιώσει όσο το δυνατόν καλύτερα την υποκειμενική μέθοδο αξιολόγησης ακρόασης χρησιμοποιώντας έναν αλγόριθμο. Η μέθοδος αυτή, όπως και όλες οι υπόλοιπες της κατηγορίας αυτής, ουσιαστικά εκμηδενίζουν το κόστος των πειραμάτων σε σχέση με τις υποκειμενικές μεθόδους και πιο συγκεκριμένα τις δοκιμές αξιολόγησης συνομιλίας και ακρόασης και κάνουν πολύ πιο εύκολη και γρήγορη την εξαγωγή αποτελεσμάτων για οποιοδήποτε τα θελήσει. Η μέθοδος μπορεί να συγκριθεί άμεσα με οποιαδήποτε άλλη που έχει πραγματοποιηθεί με τον ίδιο αλγόριθμο καθώς τα τεχνικά χαρακτηριστικά της παραμένουν όμοια. Οι διαφορές που υπάρχουν με τους υπόλοιπους αλγόριθμους είναι καθαρά σε τεχνικό επίπεδο και ουσιαστικά ο κάθε επόμενος αλγόριθμος διορθώνει αδυναμίες του προηγούμενου.

Συγκεκριμένα ο αλγόριθμος PESQ αξιολογεί σε πραγματικό χρόνο την ποιότητα ομιλίας λύνοντας το πρόβλημα αυτό του PSQM, την αξιολόγηση δηλαδή αρκετά υποβαθμισμένων βαθμολογιών λόγω του ότι ο PSQM δεν μπορούσε να υπολογίσει την χρονική καθυστέρηση

που μπορεί να είχαν τα δυο συγκρινόμενα δείγματα μεταξύ τους. Οι χρονικές καθυστερήσεις που μπορεί να έχει ένα δίκτυο είναι ποικίλες και για το λόγο αυτό η PSQM είχε στην πραγματικότητα λίγες εφαρμογές. Ο PESQ λοιπόν μπορεί να χρησιμοποιηθεί σε εφαρμογές πραγματικού χρόνου συνυπολογίζοντας τις οποιεσδήποτε καθυστερήσεις του δικτύου όπως π.χ. την απώλεια πακέτων ή την πολύ μεγάλη φυσική απόσταση της τηλεφωνικής σύνδεσης. Επίσης βελτιώνει κατά πολύ και την κλίμακα αξιολόγησης γνωμοβαθμολογίας ή αλλιώς MOS, η οποία είναι πλέον πιο σωστά διαβαθμισμένη και ακόμη πιο κοντά στα αποτελέσματα που δίνουν οι υποκειμενικές μέθοδοι αξιολόγησης. Σήμερα ο PESQ είναι ο πλέον διαδεδομένος και εμπορικά εφαρμοσμένος αλγόριθμος και χρησιμοποιείται ευρέως λόγω όλων των προηγούμενων χαρακτηριστικών παρότι στην αγορά υπάρχουν ακόμη πιο εξελιγμένοι αλγόριθμοι οι οποίοι δεν έχουν όμως έως τώρα αξιοποιηθεί επαρκώς. Η τελευταία γενιά αλγόριθμου αντικειμενικής αξιολόγησης αποτελεί ο POLQA ή Perceptual Objective Listening Quality Analysis καλύπτοντας το κενό που δημιουργούσε ο PESQ στα δίκτυα νέας γενιάς ευρείας ζώνης με υψηλή ποιότητα ήχου όπως και σε δίκτυα 3G και 4G. Ουσιαστικά πρόκειται για τον αλγόριθμο αξιολόγησης που αναμένεται να επικρατήσει τα επόμενα χρόνια και ο λόγος που δεν έχει γίνει ακόμα αυτό είναι η καθυστέρηση της μετάβασης των υφιστάμενων δικτύων στα νέας γενιάς. Οι παραπάνω αντικειμενικές μέθοδοι αποτελούν την κατηγορία των παρεμβατικών μεθόδων όπου αξιολογεί τον βαθμό υποβάθμισης του παραγόμενου σήματος σε σχέση με το αρχικό. Οι παρεμβατικές μέθοδοι βοηθούν πάρα πολύ στην περίπτωση που δεν μας ενδιαφέρει απλά το παραγόμενο αποτέλεσμα αλλά και πόση και ποια είναι η διαφορά σε σχέση με το αρχικό σήμα. Δίνουν ένα ολοκληρωμένο αποτέλεσμα εξετάζοντας όλες τις πτυχές του δικτύου από άκρο σε άκρο και έτσι μπορούμε να αξιολογήσουμε καλύτερα τους κωδικοποιητές και εν γένει τα δίκτυα αναδεικνύοντας πιθανές αδυναμίες τους. Σε περίπτωση όμως που δεν μας ενδιαφέρει μια πλήρης αξιολόγηση σε όλο το εύρος του δικτύου μπορούμε να χρησιμοποιήσουμε τις μη

παρεμβατικές αντικειμενικές μεθόδους αξιολόγησης που είναι λιγότερο πολύπλοκες κατασκευαστικά και δίνουν άμεσα αποτελέσματα με πολύ λιγότερες προϋποθέσεις ως προς την εφαρμογή των πειραμάτων. Βεβαίως τα αποτελέσματα σε αυτή την περίπτωση δεν είναι συνολικά για όλο το εύρος του δικτύου αλλά αφορούν ένα συγκεκριμένο σημείο του ή μόνο το παραγόμενο σήμα ομιλίας που τελικά λαμβάνει ο χρήστης. Οι μέθοδοι αυτές θυμίζουν στη λογική τους τις δημοσκοπικές δοκιμές υποκειμενικής αξιολόγησης με βασική διαφορά ότι τώρα εξετάζονται πέρα από τις πιθανές υποβαθμίσεις των παραγόμενων σημάτων ήχου και οι τεχνικές προδιαγραφές του δικτύου βάση των προτύπων που θέτει ο κατασκευαστής του ή ο διεξάγων το πείραμα.

Για να επαληθεύσουμε κάποια από τα παραπάνω παρατίθενται παρακάτω δυο πειράματα που διεξήγαγαν δυο εταιρίες του χώρου και δημοσίευσαν τα αποτελέσματά τους.

A) Το πρώτο πείραμα συγκρίνει τους αλγόριθμους PESQ και POLQA και διεξήχθη από την αγγλική εταιρία MALDEN. Το πείραμα χρησιμοποιεί τον κωδικοποιητή G.711 σε εργαστηριακό περιβάλλον με τα ίδια αρχεία ήχου προς εξέταση και από τους δυο αλγόριθμους. Αν και δεν παρουσιάζονται σημαντικές διαφοροποιήσεις λόγω του ότι το πείραμα εκτελείται σε συνθήκες δικτύου στενής ζώνης (NB, narrowband), όπου μπορεί να ανταποκριθεί το ίδιο καλά και ο PESQ σε σχέση με τον νεότερο POLQA, εμφανίζεται μια αξιοσημείωτη διαφορά στο δείκτη γνωμοβαθμολόγησης. Ο αλγόριθμος PESQ εμφανίζει βαθμολογία 3.58 ενώ ο POLQA εμφανίζει βαθμολογία 4.18. Παρατηρώντας το γράφημα<sup>12</sup> του PESQ βλέπουμε ότι σαφώς υπάρχει κάποια δυσλειτουργία στο τέλος της πρώτης φράσης. Αυτή η υποβάθμιση δεν υπάρχει στο γράφημα του POLQA<sup>13</sup>, στο οποίο μπορούμε να δούμε

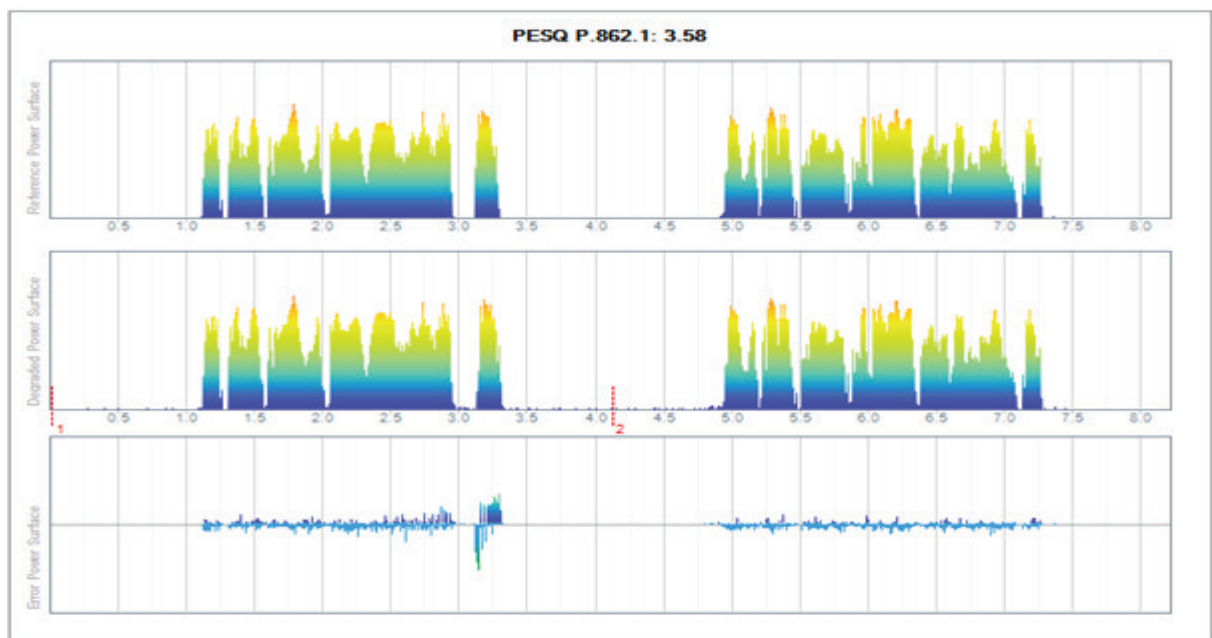
---

<sup>12</sup> Γράφημα 1

<sup>13</sup> Γράφημα 2

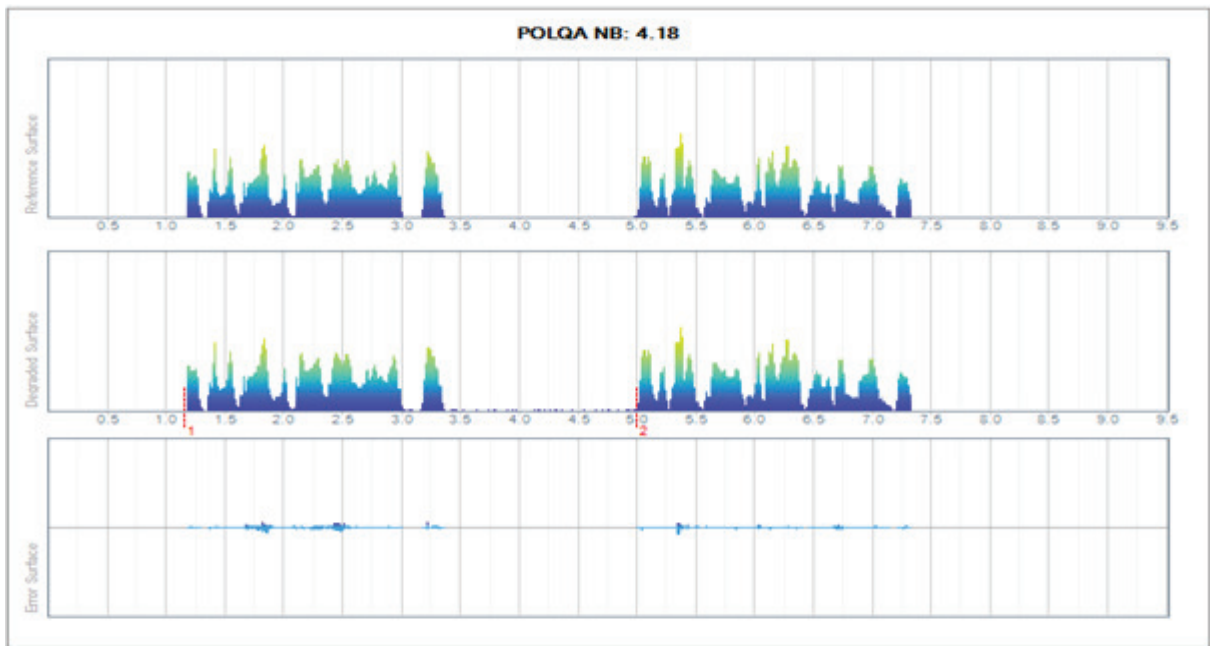
ότι υπήρξε μια διαφορά μεγέθους του αρχείου στη διάρκεια του και η οποία κράτησε για περίπου 3 δευτερόλεπτα<sup>14</sup>. Αυτό το σύντομο χρονικό διάστημα αντιστοιχεί στο διάστημα της σιωπής λίγο πριν το τέλος της πρώτης φράσης. Ο αλγόριθμος PESQ δεν μπορεί να προσδιορίσει την αλλαγή του χρόνου στο σύντομο διάστημα της σιωπής, έτσι ο τελευταίος ήχος εκλαμβάνεται από τον αλγόριθμο σαν λάθος με άμεση συνέπεια την χαμηλότερη βαθμολογία στο πείραμα παρότι όλες οι παράμετροι ήταν ίδιες. Αντίθετα ο αλγόριθμος POLQA προσδιορίζει σωστά την αλλαγή του χρόνου και η βαθμολογία που τελικά δίνει είναι η σωστή.

Το πείραμα αυτό αποδεικνύει ότι οι αντικειμενικές μέθοδοι ακριβώς επειδή πρόκειται για τεχνικά εργαλεία αξιολόγησης δεν μπορούν να αντιληφθούν παραμέτρους που στην υποκειμενική αξιολόγηση θα ήταν αυτονόητες και είναι απαραίτητη η συνεχής αναβάθμιση τους για να καλύπτει αδυναμίες ή σφάλματα των προηγούμενων εκδόσεων. Πάραταυτα η εξέλιξη κάνει τους αλγόριθμους κάθε φορά και καλύτερους φτάνοντας βήμα βήμα όλο και πιο κοντά στην αποτελεσματικότητα των υποκειμενικών μεθόδων αξιολόγησης.

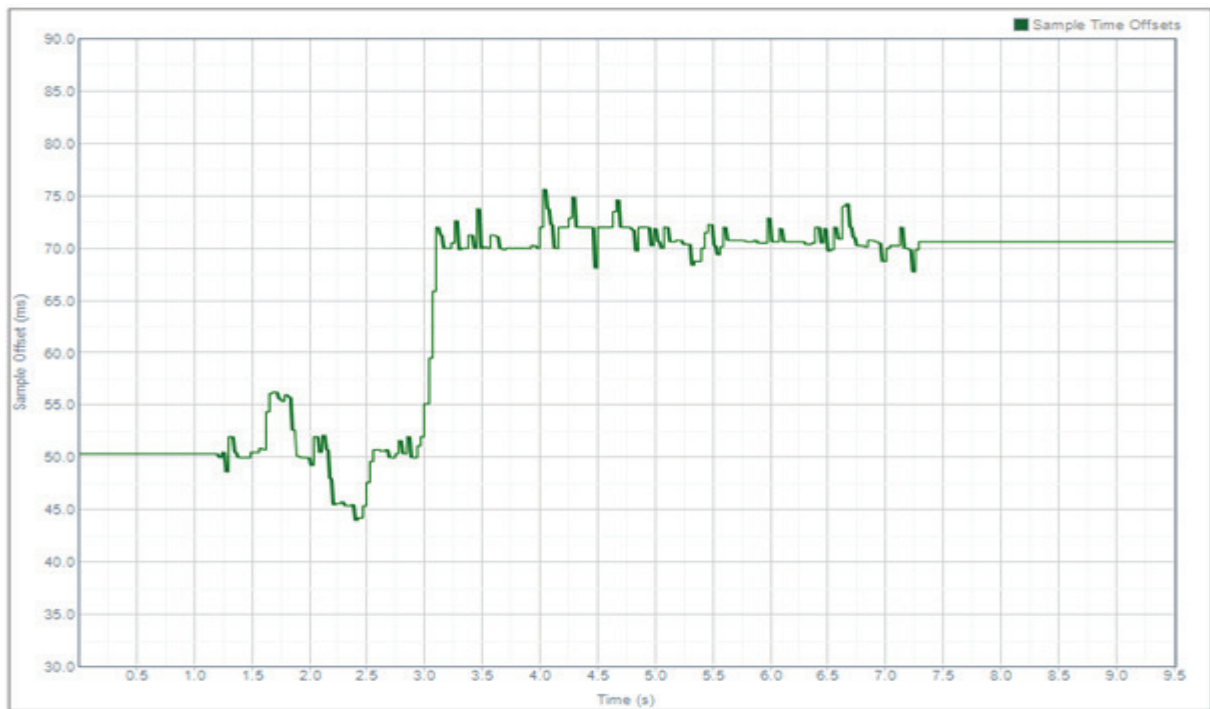


Γράφημα 1

<sup>14</sup> Γράφημα 3

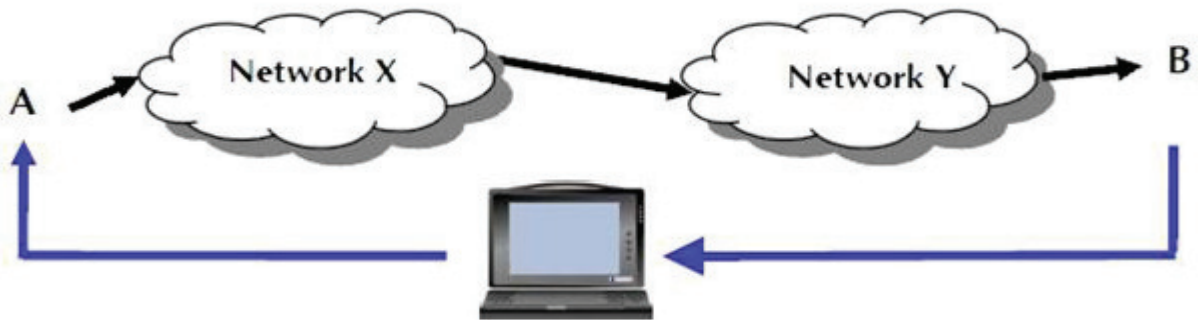


Γράφημα 2



Γράφημα 3

B) Το δεύτερο πείραμα διεξήχθη από τη εταιρία OPTICOM και συγκρίνει τις παρεμβατικές με τις μη παρεμβατικές μεθόδους αξιολόγησης και συγκεκριμένα την PESQ και την 3SQM. Οι παρεμβατικές μέθοδοι όπως ήδη έχουμε αναφέρει χρειάζονται ένα αρχικό σήμα το οποίο συγκρίνοντας το με το τελικό σήμα του ήχου βγάζουν την τελική βαθμολογία. Αυτός ο τρόπος αξιολόγησης εμφανίζεται γραφικά στην εικόνα 12.



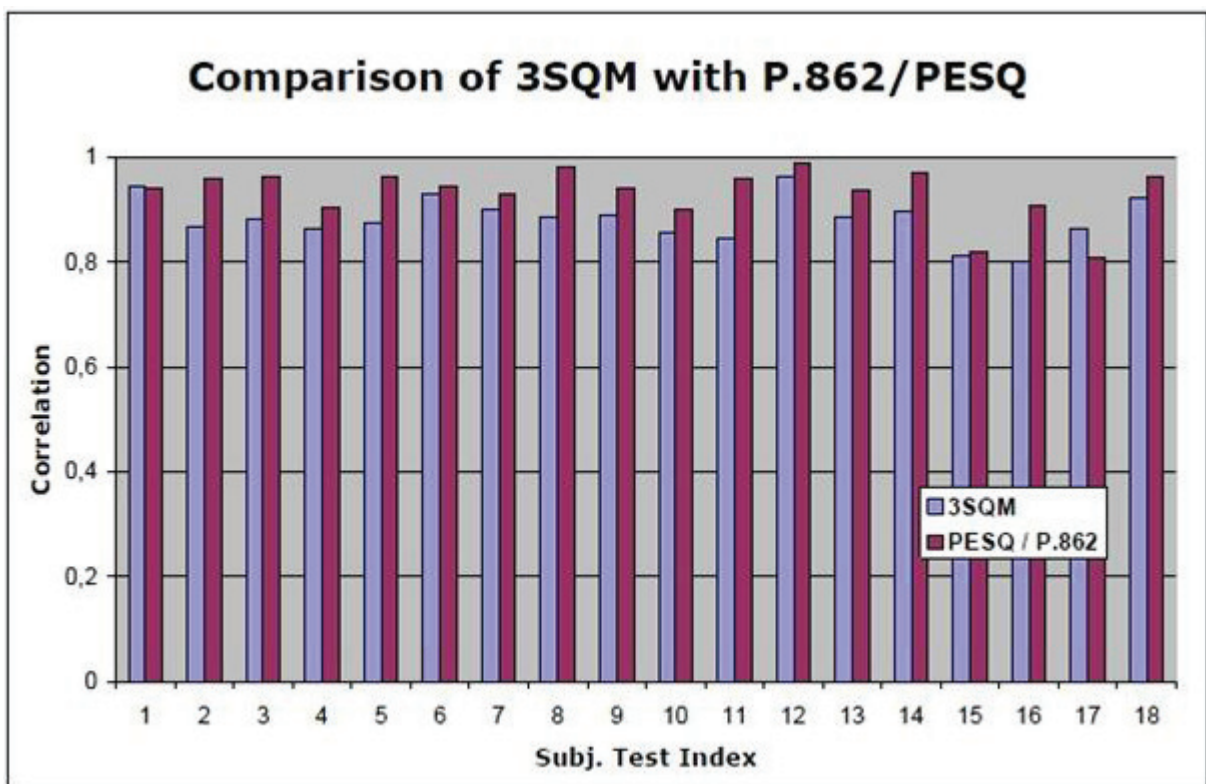
Εικόνα 13

Αντίθετα οι μη παρεμβατικές μέθοδοι δεν χρειάζονται το αρχικό σήμα ομιλίας για την βαθμολόγηση της ποιότητας του ήχου που λαμβάνεται. Αυτό που στην ουσία καταγράφουν είναι ένα και μόνο σημείο το οποίο χρησιμοποιείται σαν δείκτης της παρεχόμενης ποιότητας στο δίκτυο σε πραγματικό χρόνο. Γραφική παράσταση της εν λόγω μεθόδου παρουσιάζεται στην εικόνα 13.



Εικόνα 14

Τα αποτελέσματα<sup>15</sup> του πειράματος λοιπόν παρουσιάζουν την απόδοση του αλγόριθμου PESQ σε σχέση με τον αλγόριθμο 3SQM. Οι συσχετισμοί μεταξύ των αποτελεσμάτων από την 3SQM και την PESQ εμφανίζονται ανά βάση δεδομένων για την καθεμία ξεχωριστά. Είναι σημαντικό ότι η 3SQM παρουσιάζει αξιοσημείωτες επιδόσεις και πολλές φορές φτάνει αυτές της PESQ. Έχοντας υπ' όψιν την πολύ μεγαλύτερη ευελιξία των μη παρεμβατικών μεθόδων και με δεδομένο τα αρκετά καλά αποτελέσματα βαθμολόγησης μπορούμε να θεωρήσουμε ότι πρόκειται για μια πολύ καλή λύση συνεχούς παροχής δείκτη ποιότητας ενός δικτύου. Έτσι σε περίπτωση που παρατηρηθεί κάποια δυσλειτουργία ή μη συνήθης μέτρηση ποιότητας μπορούμε να χρησιμοποιήσουμε άμεσα έναν αλγόριθμο παρεμβατικής μεθόδου όπως ο PESQ για την εξαγωγή ακριβών αποτελεσμάτων.



Γράφημα 4

<sup>15</sup> Γράφημα 4



## ***BIBΛΙΟΓΡΑΦΙΑ***

- 📖 ITU-T Recommendation P.562, Analysis and interpretation of INMD voice-service Measurements, 2004
- 📖 ITU-T Recommendation P.563, Single-ended method for objective speech quality Assessment in narrow-band telephony applications, 2004
- 📖 ITU-T Recommendation P.800, Methods for subjective determination of transmission quality, 1996
- 📖 ITU-T Recommendation P.800.1, Mean Opinion Score (MOS) terminology, 2006
- 📖 ITU-T Recommendation P.861, Objective Quality Measurement of Telephone-Band(300-3400 Hz) Speech Codecs, 1998
- 📖 ITU-T Recommendation P.862, Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band Telephone networks and speech codecs, 2001.
- 📖 ITU-T Recommendation P.863, Perceptual Objective Listening Quality Assessment, 2011
- 📖 OPTICOM GmbH P.563 - Single Side Speech Quality Measure, 2005  
<http://www.opticom.de>
- 📖 ITU-T Study Group 12. E-model Calculation Tool, 2012.  
<http://www.itu.int/ITU-T/studygroups/com12/emodelv1/index.htm>
- 📖 ASCOM -The PESQ Algorithm as the Solution for Speech Quality Evaluation on 2.5G and 3G Networks
- 📖 Speech Coding: <http://www.cslu.ogi.edu/HLTsurvey/ch10node4.html>
- 📖 G.7xx: Audio (Voice) Compression Protocols (CODEC) (G.711, G.721, G.722, G.726, G.727, G.728, G.729): <http://www.javvin.com/protocolG7xx.html>

- 📖 Speech Codecs: Pros & Cons: <http://speechcodecs.wordpress.com/>
- 📖 Understanding Codecs: Complexity, Hardware Support, MOS and Negotiation:  
[http://www.cisco.com/en/US/tech/tk1077/technologies\\_tech\\_note09186a00800b6710.shtml](http://www.cisco.com/en/US/tech/tk1077/technologies_tech_note09186a00800b6710.shtml)
- 📖 NIQA: [http://arptel.com/products\\_vq\\_niqa.html](http://arptel.com/products_vq_niqa.html)
- 📖 NIQA: <http://www.sevana.fi/non-intrusive-voice-quality-testing-software.php>
- 📖 Timothy A. Hall : "Objective speech quality measures for Internet telephony", *Proc. SPIE 4522, Voice Over IP (VoIP) Technology*, 128 (July 25, 2001)
- 📖 Voice Quality Measurement: [http://www.technology-training.co.uk/voicequalitymeasurement\\_39.php](http://www.technology-training.co.uk/voicequalitymeasurement_39.php)
- 📖 The E-Model : <http://www.itu.int/ITU-T/studygroups/com12/emodelv1/tut.htm>
- 📖 The E Model / Predicting the perceived quality of service with the R Value :  
<http://www.datarave.net/zfh/tag/e-model/>
- 📖 Omair S. Malik, Quality of Service Analysis for Audio over Cellular Voice Networks and Cellular Wireless Wide Area Networks, Massachusetts Institute of Technology, 2007
- 📖 NetIQ, Assessing VoIP Call Quality Using the E-model
- 📖 ITU-T Recommendation P.862, Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs
- 📖 3SQM: <http://www.opticom.de/technology/3sqm.php>
- 📖 P.563: [http://www.opticom.de/download/02SpecSheet\\_P563\\_06-03-03.pdf](http://www.opticom.de/download/02SpecSheet_P563_06-03-03.pdf)
- 📖 Opticom, Advanced Non-Intrusive voice quality testing
- 📖 Malden Electronics, POLQA Gets the Timing Right:  
<http://www.malden.co.uk/2012/10/polqa-gets-the-timing-right/>