



**ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ**

Ψηφιακή Επεξεργασία Φωνής

Διάλεξη: Προσαρμόσιμο Αρμονικό Μοντέλο

Παρουσίαση: Gilles Degottex

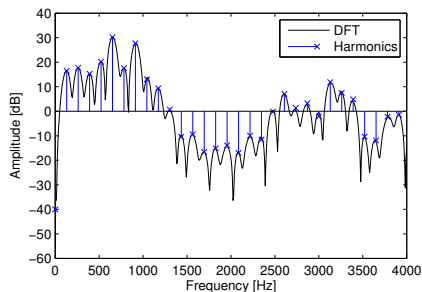
Στυλιανού Ιωάννης
Τμήμα Επιστήμης Υπολογιστών

A Full-Band Adaptive Harmonic Representation of Speech

Gilles Degottex and Yannis Stylianou

University of Crete - FORTH - Swiss National Science Foundation

The Sinusoidal and Harmonic Models



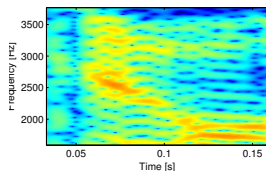
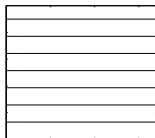
Can fit any monophonic signal, we use it for speech
The sinusoids can be harmonic, quasi-harmonic, or adaptive ...

Time-Frequency Representations

DFT $s(t) = \sum_{k=0}^K a_k \cdot e^{j\phi_k(t)}$

$$\phi_k(t) = k \cdot (2\pi/K) \cdot t$$

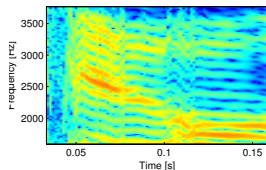
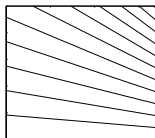
Constant frequency basis



FChT¹ $s(t) = \sum_{k=0}^K a_k \cdot e^{j\phi_k(t)}$

$$\phi_k(t) = k \cdot (2\pi/K + \alpha \cdot t) \cdot t$$

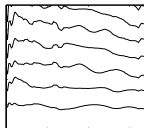
Linear frequency basis



¹M. Kepesi and L. Weruaga, *Adaptive Chirp-based time-frequency analysis of speech signals*, Speech communication, 2006.

The Adaptive Quasi-Harmonic + Noise Model (aQHNM)¹

- We can adapt the frequency basis to follow the frequency tracks
 - ⇒ Adaptive Quasi-Harmonic Model (aQHM)¹
$$\phi_k(t) = \frac{2\pi}{f_s} \int_0^t f_k(\tau) d\tau$$
- For speech representation in the high frequencies
 - ⇒ Amplitude modulated noise (aQHNM)²



¹Y. Pantazis, O. Rosenc and Y. Stylianou, *Adaptive AM-FM Signal Decomposition With Application to Speech Analysis*, IEEE Trans. on Audio, Speech, and Language Processing, 2010.

²Y. Pantazis, G. Tzedakis, O. Rosenc, Y. Stylianou, *Analysis/Synthesis of Speech based on an Adaptive Quasi-Harmonic plus Noise Model*, ICASSP, 2010.

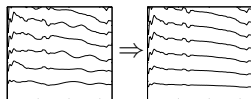
The new ideas

- 1) From FChT, harmonics exist in high frequencies
⇒ Use a full-band representation

The new ideas

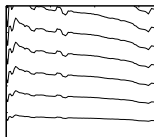
- 1) From FChT, harmonics exist in high frequencies
⇒ Use a full-band representation
- 2) Quasi-harmonicity can be useful for analysis but maybe not necessary for encoding/decoding
⇒ Use the strict harmonicity and keep the adaptivity

aQHNM ⇒ aHM



The Adaptive Harmonic Model (aHM)

$$\text{aHM } s(t) = \sum_{k=-K}^K a_k(t) \cdot e^{j\phi_k(t)}$$
$$\phi_k(t) = k \cdot \frac{2\pi}{f_s} \int_0^t f_0(\tau) d\tau$$



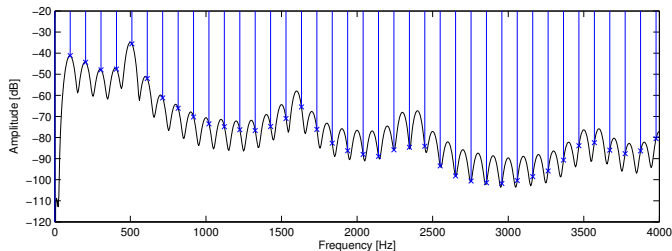
$a_k(t)$ Amplitude and phase (complex-valued function)
Interpolated from $a_k^{t_i}$ at time t_i

$f_0(t)$ Fundamental frequency
Interpolated from $f_0^{t_i}$ at time t_i

Parameters at a time t_i : $\{f_0^{t_i}, a_k^{t_i}\} \quad k \in \{0, \dots, K_i\}$

The problem of estimation for full-band representation

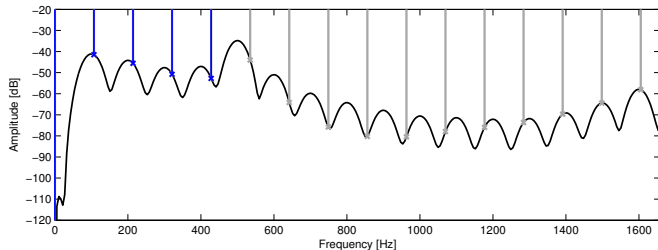
A small f_0 error propagates by multiplication: $f_k = k \cdot f_0$



Question How to estimate harmonics up to Nyquist ?

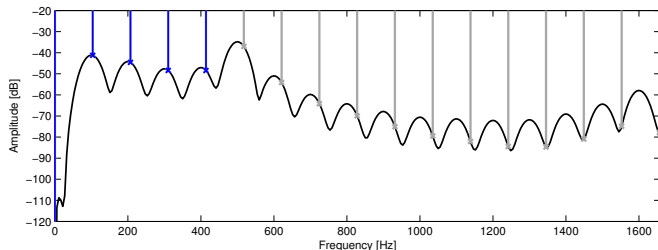
The Adaptive Iterative Refinement (AIR)

Assume first the f_0 error is small for low harmonics



The Adaptive Iterative Refinement (AIR)

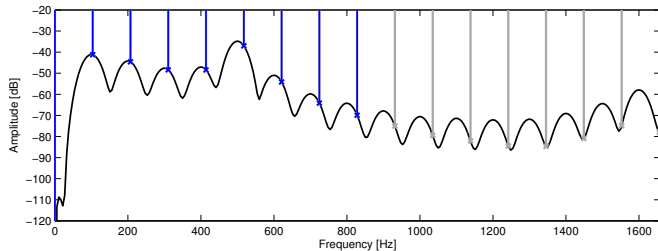
Then the frequency correction mechanism of QHM¹ can be used



¹Y. Pantazis, O. Rosec and Y. Stylianou, *Iterative Estimation of Sinusoidal Signal Parameters*, IEEE Signal Processing Letters, 2010.

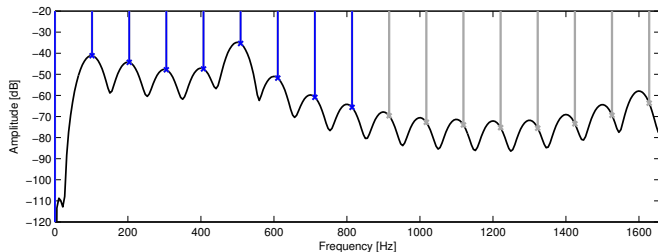
The Adaptive Iterative Refinement (AIR)

We can therefore increase the harmonic level



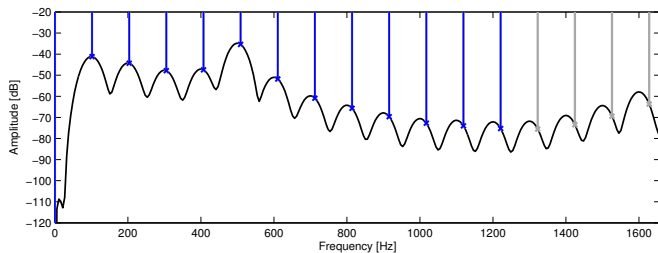
The Adaptive Iterative Refinement (AIR)

Correct the frequencies



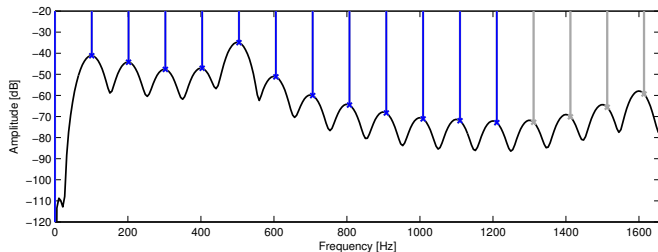
The Adaptive Iterative Refinement (AIR)

Increase the harmonic level



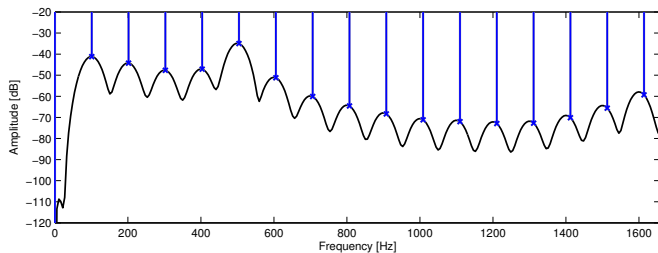
The Adaptive Iterative Refinement (AIR)

Correct the frequencies



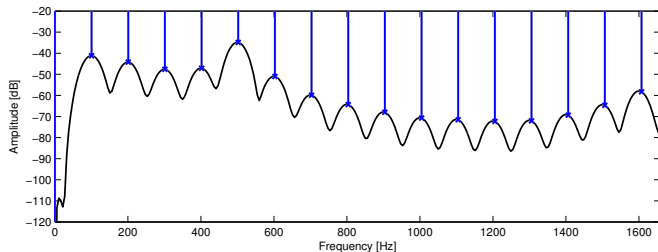
The Adaptive Iterative Refinement (AIR)

Increase the harmonic level

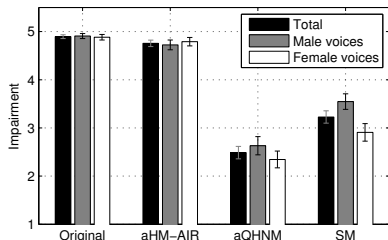


The Adaptive Iterative Refinement (AIR)

Correct the frequencies



Evaluation: Listening test



- 6 languages to represent voice variability
- Female and male voices for each language \Rightarrow 12 sounds
- 20 listeners answered

Conclusions

- + Perceived quality aHM-AIR is almost perfect
 - Compared to SM: stable frequency tracks in aHM
 - Compared to aQHNM: no noise model in aHM, also more stable

Conclusions

Points to remember

- Adaptive Harmonic Model (aHM)
 - Frequency tracks adapted to the f_0 curve
 - Simple harmonicity

Conclusions

Points to remember

- Adaptive Harmonic Model (aHM)
 - Frequency tracks adapted to the f_0 curve
 - Simple harmonicity
- Dedicated algorithm, Adaptive Iterative Refinement (AIR), to localize the harmonic structures in the high frequencies

Conclusions

Points to remember

- Adaptive Harmonic Model (**aHM**)
 - Frequency tracks adapted to the f_0 curve
 - Simple harmonicity
- Dedicated algorithm, Adaptive Iterative Refinement (**AIR**), to localize the harmonic structures in the high frequencies
- Quasi-perfect perceived quality according to a listening test

Conclusions

Points to remember

- Adaptive Harmonic Model (aHM)
 - Frequency tracks adapted to the f_0 curve
 - Simple harmonicity
- Dedicated algorithm, Adaptive Iterative Refinement (AIR), to localize the harmonic structures in the high frequencies
- Quasi-perfect perceived quality according to a listening test
- Less parameters than aQHNM and SM

Σας ευχαριστώ για την προσοχή σας

Τέλος Ενότητας



Ευρωπαϊκή Ένωση
Πρωτόκολλο Συνεταιρισμού



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «**Εκπαίδευση και Δια Βίου Μάθηση**» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Σημειώματα

Σημείωμα αδειοδότησης

- Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση, Όχι Παράγωγο Έργο 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».

[1] <http://creativecommons.org/licenses/by-nc-nd/4.0/>



- Ως **Μη Εμπορική** ορίζεται η χρήση:
 - που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
 - που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
 - που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο
- Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Κρήτης, Στυλιανού Ιωάννης. «Ψηφιακή Επεξεργασία Φωνής. Προσαρμόσιμο Αρμονικό Μοντέλο». Έκδοση: 1.0. Ηράκλειο/ Ρέθυμνο 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: <http://www.csd.uoc.gr/~hy578>