# BIOMET®TOOLS: FROM MODELLING AND SIMULATION TO PRODUCT DESIGN AND DEVELOPMENT

Pedro Gómez, Victoria Rodellar, Víctor Nieto, Rafael Martínez and Agustín Álvarez
Neuromorphic Speech Processing Laboratory, Centro de Tecnología Biomédica, Universidad Politécnica de Madrid
Campus de Montegancedo, s/n 28660 Boadilla del Monte, Madrid, Spain
Bartolomé Scola, Carlos Ramírez, Daniel Poletti, Mario Fernández
ORL and ENT Services, Hospital Universitario Gregorio Marañón, Hospital del Henares, Madrid
e-mail: pedro.gomez@ctb.upm.es

ABSTRACT: BioMet®Tools is a set of software applications developed for the biometrical characterization of voice in different fields as voice quality evaluation in laryngology, speech therapy and rehabilitation, education of the singing voice, forensic voice analysis in court, emotional detection in voice, secure access to facilities and services, etc. Initially it was conceived as plain research code to estimate the glottal source from voice and obtain the biomechanical parameters of the vocal folds from the spectral density of the estimate. This code grew to what is now the Glottex®Engine package (G®E). Further demands from users in medical and forensic fields instantiated the development of different Graphic User Interfaces (GUI's) to encapsulate user interaction with the G®E. This required the personalized design of different GUI's handling the same G®E. In this way development costs and time could be saved. The development model is described in detail leading to commercial production and distribution. Study cases from its application to the field of laryngology and speech therapy are given and discussed.

## 1. INTRODUCTION

The present paper is intended to give an overview on product design and development from an end-user-driven application which started simply as computer software to study a specific phenomenon: the glottal source and its associated mucosal wave correlate [1]. The glottal source may be seen as the pressure build-up in the glottis just above the vocal folds in the laryngeal cavity. It is the result of the phonation cycle, seen as a sequence of openings and closings of the vocal folds under the influence of lung pressure and vocal fold viscoelasticity and air dynamics [2]. The glottal source is expected to follow closely the theoretical pattern proposed by G. Liljencrants and G. Fant [3] known as the L-F pattern given in Figure 1.

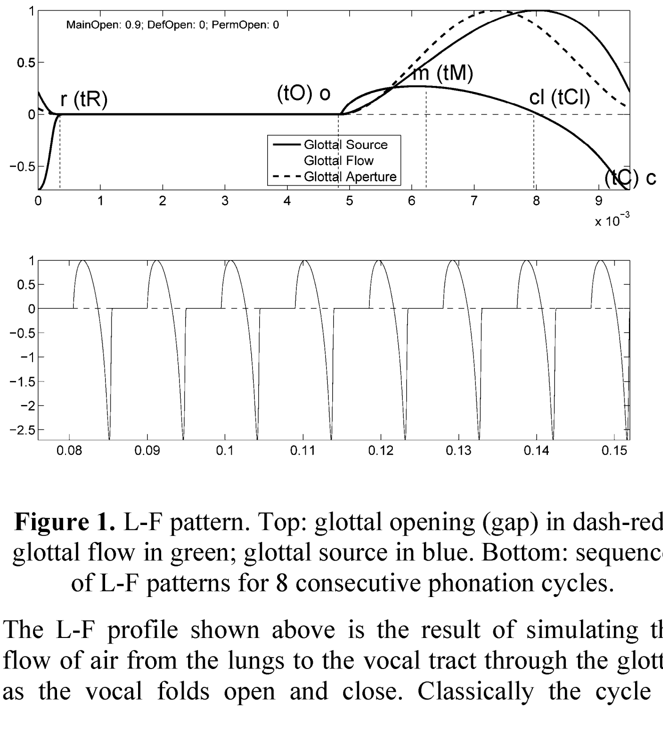


**Figure 1.** L-F pattern. Top: glottal opening (gap) in dash-red; glottal flow in green; glottal source in blue. Bottom: sequence of L-F patterns for 8 consecutive phonation cycles.

The L-F profile shown above is the result of simulating the flow of air from the lungs to the vocal tract through the glottis as the vocal folds open and close. Classically the cycle is considered to start at the opening instant (tO), nevertheless, as this instant sometimes is rather inaccurate, the closing instant (tC) is preferred. The sudden stop of the airflow in the larynx, pharynx and vocal tract results in a sudden drop of the dynamic pressure from 0 to a minimum (at t=0). After a time interval lasting from a fraction of a millisecond to 1-2 ms, the dynamic pressure recovers to its quiescent value (0). This instant is signaled as tR (recovery instant). During the remnant part of the closed phase the vocal folds are supposedly in contact and no airflow is allowed through the glottis. The dynamic pressure remains in its resting value (0). The vocal folds start opening (the equivalent light seen through the glottis is called the *gap* in dash red), and this results in a pressure build-up towards a maximum (tM) where the airflow is in its steepest ascent (green line). When the opening reaches the maximum value the pressure is dropping to the resting value again, but as the vocal folds come closer (adduction) the pressure drop is larger, crossing the resting value at tCl, and falling to a minimum when both vocal folds produce a complete flow stop (tC). The phonation and glottal cycles repeat the same pattern once and again. From what has been said, it may seem clear that the specific profiles of the recovery, closed, open and closing phases will reveal important details of the system biomechanics. A good reconstruction of the glottal source is of most relevance to ensure proper estimates of the system biomechanics. For such, a careful removal of the vocal tract by system inversion is necessary. The interested reader is referred to [1] for a complete explanation.

## 2. THE G®E TECHNOLOGY

The technology encapsulated in the Glottex®Engine is based on the detection of the glottal source from the inversion of the vocal tract and the removal of its influence from voice. A good example of the glottal source reconstruction is shown in Figure 2. The example is taken from a male subject with normophonic voice, non-smoker, pathology-free condition assessed by objective endoscopy. The true and pseudo-recovery and opening instants are given as tR1, tR2, tO1 and tO2.

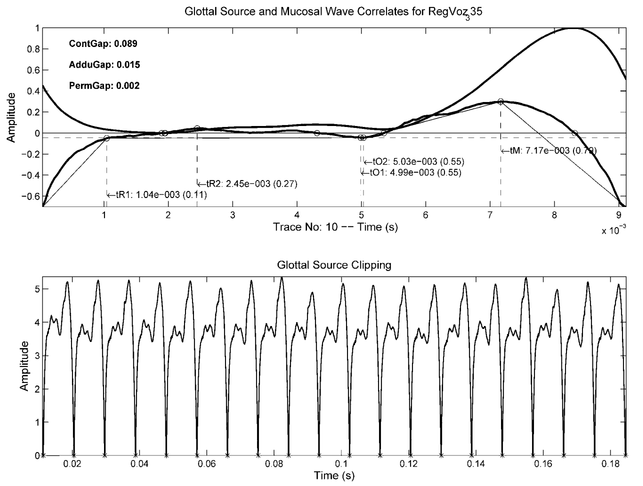


**Figure 2.** Typical glottal source. Top: a glottal cycle spanning from a closing instant to the next closing instant. Bottom: Sequence of glottal cycles in an interval of 183 ms.

The recovery and contact estimates are quite realistic and resemble the simulated pattern in Figure 1. A set of 65 parameters including distortion, cepstral, spectral, biomechanical, temporal, contact and tremor are obtained from the glottal source following the methodology in Figure 3.

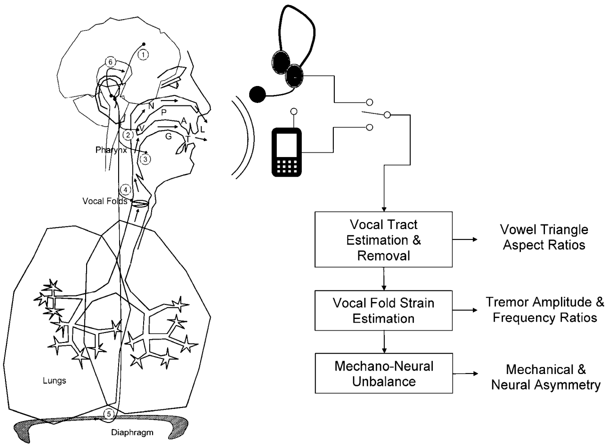
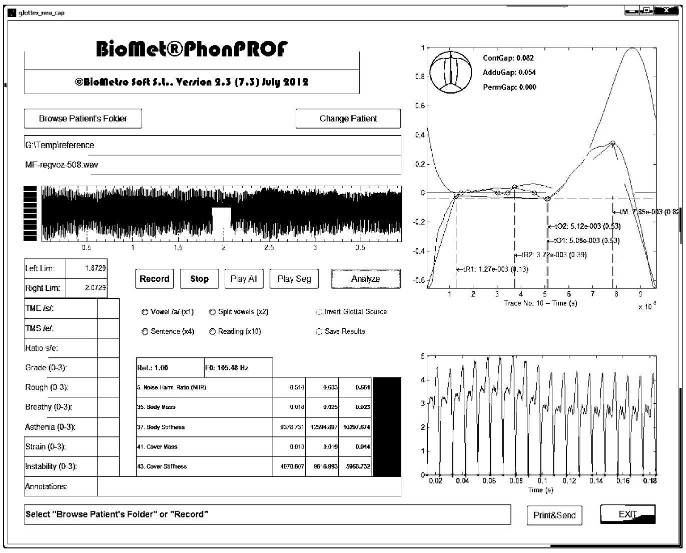

Pharynx
Vocal Folds
Lungs
Diaphragm
N
P
G
L
Vocal Tract Estimation & Removal
Vowel Triangle Aspect Ratios
Vocal Fold Strain Estimation
Tremor Amplitude & Frequency Ratios
Mechano-Neural Unbalance
Mechanical & Neural Asymmetry

**Figure 3.** Model inversion to estimate vocal tract, biomechanical and neurological parameters from voice.

This extraction methodology is referred to as the Glottal Engine technology or G®E and is compiled as a C++ package generated from MATLAB®.

## 3. THE MODEL BEHIND BIOMET®TOOLS

The G®E package may be adapted to different purposes by the intermediation of different Graphic User Interfaces. One such interface (BioMet®Phon) is shown Figure 4 for use in Voice Quality Analysis by Laryngologists or Speech Therapists. The GUI is rather simple to use: a new voice recording, analysis and full automatic report in Adobe®pdf, and an Excel® document with the statistical distributions of the estimated parameters may be generated in less than 10 s with three button clicks. The GUI allows the handling of a small patient's database. Once a patient is selected either a new recording may be obtained and analyzed or an old one may be processed. A sketch of the glottal source is presented in the upper right window of Figure 4. A set of five selected parameters are presented in comparative windows (mid bottom) showing normality limits and ticketing the results as green (within normality) or red (out of normality). This code allows a fast semantic interpretation by the laryngologist or speech therapist.



©BioMetro Soft S.L., Version 2.3 (7.3) July 2012

G:\Temp\reference

MF-regvoz-508.wav

| | |
|---|---|
| Left Lim: | 1.8729 |
| Right Lim: | 2.0729 |

○ Vowel /a/ (x1) ○ Split vowels (x2) ○ Invert Glottal Source

○ Sentence (x4) ○ Reading (x10) ○ Save Results

TME /s/:

TMS /e/:

Ratio s/e:

Grade (0-3):

Rough (0-3):

Breathy (0-3):

Asthenia (0-3):

Strain (0-3):

Instability (0-3):

Annotations:

| Rel.: 1.00 | F0: 105.48 Hz | | |
|---|---|---|---|
| 5. Noise-Harm. Ratio (NHR) | 0.510 | 0.633 | 0.551 |
| 35. Body Mass | 0.018 | 0.025 | 0.023 |
| 37. Body Stiffness | 9378.731 | 12594.897 | 10297.674 |
| 41. Cover Mass | 0.010 | 0.019 | 0.014 |
| 43. Cover Stiffness | 4070.607 | 9616.993 | 5958.732 |

Select "Browse Patient's Folder" or "Record"

**Figure 4.** GUI of BioMet®Phon.

BioMet®Phon produces two kinds of results: visual documents estimating the power spectral density of voice and the glottal source for specific comparisons as in Figure 5 or a global report as the one in Figure 6, both as Adobe®pdf documents. Other GUI's based on G®E are designed for Forensic Voice Analysis (BioMet®Fore) or the study of the Singing Voice (BioMet®Sing). This suite is known as BioMet®Tools.

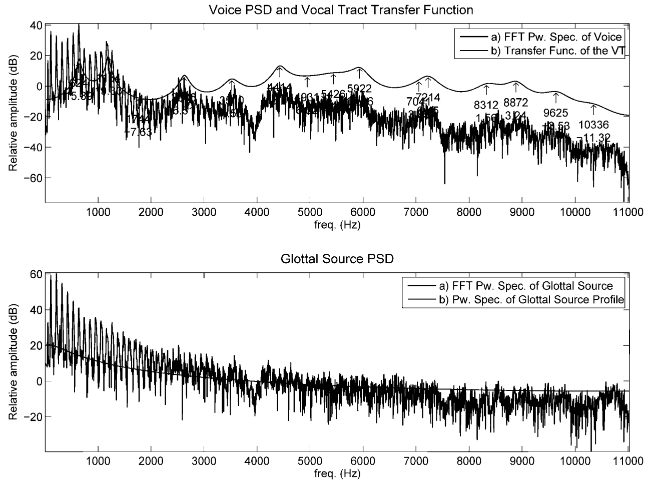
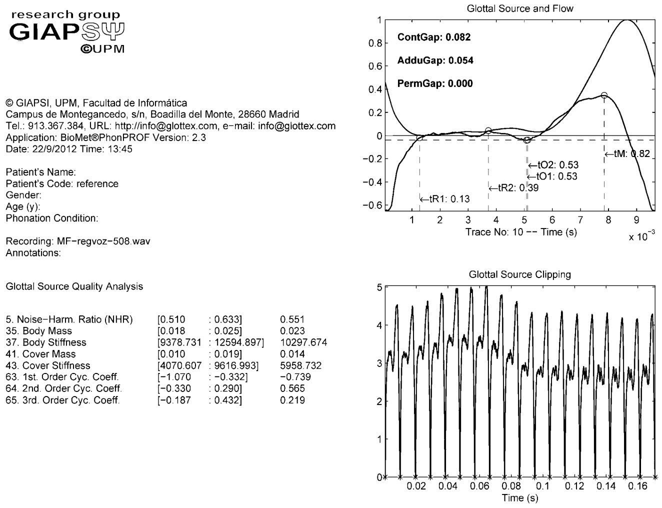


**Figure 5.** Top: Power spectral density of voice in blue and vocal tract resonances in red. Bottom: Power spectral density of the glottal source in blue and bottom source profile in red.


research group
GIAPSΨ
©UPM
Glottal Source and Flow
ContGap: 0.082
AdduGap: 0.054
PermGap: 0.000
tM: 0.82
tO2: 0.53
tO1: 0.53
tR2: 0.39
tR1: 0.13
Trace No: 10 -- Time (s)
x 10^-3
Glottal Source Clipping
Time (s)

© GIAPSI, UPM, Facultad de Informática
Campus de Montegancedo, s/n, Boadilla del Monte, 28660 Madrid
Tel.: 913.367.384, URL: http://info@glottex.com, e-mail: info@glottex.com
Application: BioMet®PhonPROF Version: 2.3
Date: 22/9/2012 Time: 13:45

Patient's Name:
Patient's Code: reference
Gender:
Age (y):
Phonation Condition:

Recording: MF−regvoz−508.wav
Annotations:

Glottal Source Quality Analysis

| | | | |
|---|---|---|---|
| 5. Noise−Harm. Ratio (NHR) | [0.510 | : 0.633] | 0.551 |
| 35. Body Mass | [0.018 | : 0.025] | 0.023 |
| 37. Body Stiffness | [9378.731 | : 12594.897] | 10297.674 |
| 41. Cover Mass | [0.010 | : 0.019] | 0.014 |
| 43. Cover Stiffness | [4070.607 | : 9616.993] | 5958.732 |
| 63. 1st. Order Cyc. Coeff. | [−1.070 | : −0.332] | −0.739 |
| 64. 2nd. Order Cyc. Coeff. | [−0.330 | : 0.290] | 0.565 |
| 65. 3rd. Order Cyc. Coeff. | [−0.187 | : 0.432] | 0.219 |

**Figure 6.** Example of report generated as an Adobe®pdf document.

# 4. PRE-POST-TREATMENT RESULTS

In what follows a typical study case will presented showing how BioMet®Phon may be used in assessing voice quality improvement after treatment. A specific case of pre-treatment compared with three post-treatment inspections is presented and discussed. It corresponds to a female patient 65 years-old who suffered from post-Thyroidectomy Vocal Fold Recurrent Paralysis (pTVFRP). The treatment consisted in infiltration of fat from the patient in the vocal fold. The patient's voice was examined once before the intervention (pre: March) and three times after the intervention (post1: May; post2: September; and post3: November) all over 2011. The 8 most relevant parameters for dysphonic voice evaluation were selected from the set of 65 ones after each examination and are listed in Table 1. These same data may be seen plotted in Figure 7.

**Table 1.** Results of pre- and post-treatment for a specific case (pTVFRP) on a set of selected parameters.

| Parameter | Pre | Post1 | Post2 | Post3 |
|---|---|---|---|---|
| 2-Jitter (%) | 2.8 | 5.4 | 0.6 | 0.6 |
| 3-Shimmer (%) | 10.5 | 3.3 | 1.5 | 1.0 |
| 38-Body M. Unb. (%) | 4 | 21 | <1 | <1 |
| 40-Body S. Unb. (%) | 10 | 30 | 1 | 1 |
| 41-Cover M. (mg) | 26 | 8 | 8 | 6 |
| 43-Cover S. (g.s$^{-2}$) | 91,746 | 24,228 | 14,175 | 11,808 |
| 44-Cover M. Unb. (%) | 47 | 14 | 2 | 1 |
| 46-Cover S. Unb. (%) | 43 | 26 | 3 | 3 |

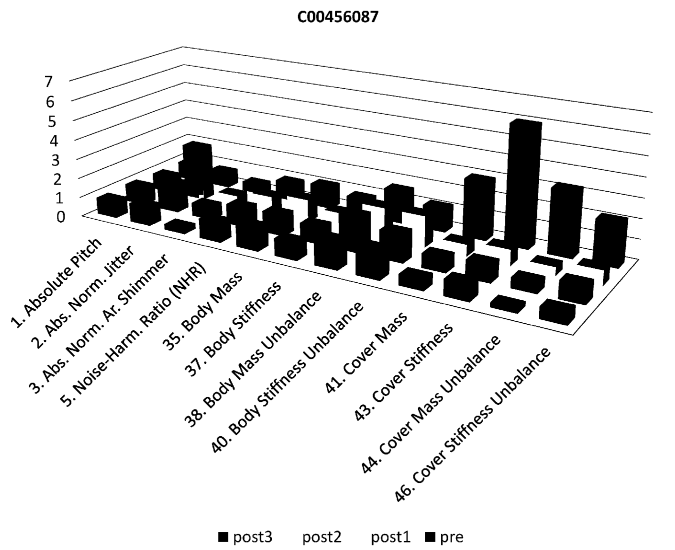


**Figure 7.** Results of pre- and post-treatment for a specific case of pTVFRP normalized on the reference female set medians.

Classically *2-Jitter* and *3-Shimmer* are parameters used very often in voice quality evaluation, as they are known to be well correlated with dysphonic voice [4]. Nevertheless these parameters lack structural semantics, as they do not allow producing hypotheses on possible etiological circumstances. Biomechanical parameters as the subset left (38: Body Mass Unbalance, 40: Body Stiffness Unbalance, 41: Cover Mass, 43: Cover Stiffness, 44: Cover Mass Unbalance and 45: Cover Stiffness Unbalance) allow casting hypotheses on possible etiological implications based on the differential structural correlates conveyed by them. For instance, it seems clear that jitter (2) correlates more with fold body parameters (38, 40), whereas shimmer (3) is more related to cover parameters (41, 43, 44, 46). As jitter and body parameters suffer an increment after intervention (in *post1* relative to *pre*) contrary to shimmer and cover parameters, it seems that the intervention affected the fold body in a different way than to cover. This simple reflection puts into consideration that the semantics of the biomechanical parameters is considerably larger than that of classical distortion ones.

# 5. VALIDATION AND DISCUSSION

Ultimately the objective of an application to evaluate voice quality is to produce accurate results in detecting dysphonic voice from normal. Therefore a validation of the application should be provided. This is achieved using a database of 200 subjects collected at Hospital Universitario Gregorio Marañón divided into two subsets of 100 subjects equally balanced by gender, and these on their turn comprising half normophonic and half dysphonic subjects. Therefore the set used in the study consisted in 50+50+50+50 subjects balanced by gender and voicing condition. The age span covered from 20 to 60 years, the medians in 35 for male and 34 for females. Sustained phonation emissions of vowel /a/ were recorded in three different sessions. Samples of 200 ms of each emission were used in the extraction of a set of 65 parameters for each phonation cycle. Estimations of medians (Q2), first (Q1) and third quartiles (Q3) were used as distribution descriptors for each emission. Medians from each emission were used in the study, to evaluate the probability of a given patient observation $\mathbf{x_q}$ being associated to the respective gender normophonic set:

$$\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_m}) = \frac{1}{(2\pi)^{P/2}\left|\mathbf{C_m}\right|^{1/2}} \iiint_{(-\infty,\mathbf{x_q})} e^{-1/2(\zeta-\chi_m)^T \mathbf{C}_m^{-1}(\zeta-\chi_m)} \mathbf{d}\zeta$$
$$\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_f}) = \frac{1}{(2\pi)^{P/2}\left|\mathbf{C_f}\right|^{1/2}} \iiint_{(-\infty,\mathbf{x_q})} e^{-1/2(\zeta-\chi_f)^T \mathbf{C}_f^{-1}(\zeta-\chi_f)} \mathbf{d}\zeta \qquad (1)$$

where $\mathbf{x_q}$ is the observations vector of dimension *P* for subject *q*, and $\mathbf{\Gamma_m}$={$\mathbf{C_m}$, $\chi_m$} and $\mathbf{\Gamma_f}$={$\mathbf{C_f}$, $\chi_f$} are the respective Gaussian models for the male (m) and female (f) datasets, with the mean vectors $\chi_m$ and $\chi_f$ and the Covariance Matrices $\mathbf{C_m}$ and $\mathbf{C_f}$ to be estimated on each gender set. The likelihood of each subject given a label *v* as normophonic (*n*) or dysphonic (*d*) relative to his/her gender set will then compared to a certain threshold *θ*:

$$\lambda_m(\mathbf{x_q}) = \log\frac{\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_m})}{1-\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_m})}; \quad \nu_m(\mathbf{x_q}) = \begin{cases} n & \text{if} \quad \lambda_m \geq \theta \\ d & \text{if} \quad \lambda_m < \theta \end{cases}$$
$$\lambda_f(\mathbf{x_q}) = \log\frac{\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_f})}{1-\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_f})}; \quad \nu_f(\mathbf{x_q}) = \begin{cases} n & \text{if} \quad \lambda_f \geq \theta \\ d & \text{if} \quad \lambda_f < \theta \end{cases} \qquad (2)$$

The database was processed using a ten-time cross-validation procedure removing 5 subjects each time out of 50 within a ten-time scale, thus producing 1000 scores per gender set. The results are plotted in Figure 8 and Figure 9 for each respective gender set as Tippet plots, ROC (Receiver Operator Characteristic) and DET (Detection-Error Trade-off) curves [5]. The results show fairly similar detection capabilities for both genders. Tippett plots (upper right templates) showing the evolution of false positive and negative detections accordingly with the selection of the threshold are especially relevant. If the threshold θ to admit a normophonic voice is very low, most of the population will be labelled as normophonics (the number of False Normophonics will be large), and all normophonics will be correctly labelled (no False Dysphonics). Conversely high values of θ will present the opposite situation. The DET curves (lower right) if scaled in logarithmic axes offer a clear view of the Equal Error Rate point (EER), which is the point of the curve where the rate of False Positives and Negatives equal. This can be taken as a merit factor, which is around 2.7% for the male set and 3.2% for the female set. These curves allow designing different detection scenarios. For instance, to reduce the rate of False Negatives to 1% in the male set an assumption of 15% False Positives should be admitted. As the emission of a False Negative in the detection of dysphonic voice is far more sensitive that the production of a False Positive, it should be admitted that around 1 out of 6 subjects with normal voice would be detected as possible dysphonic not being so.

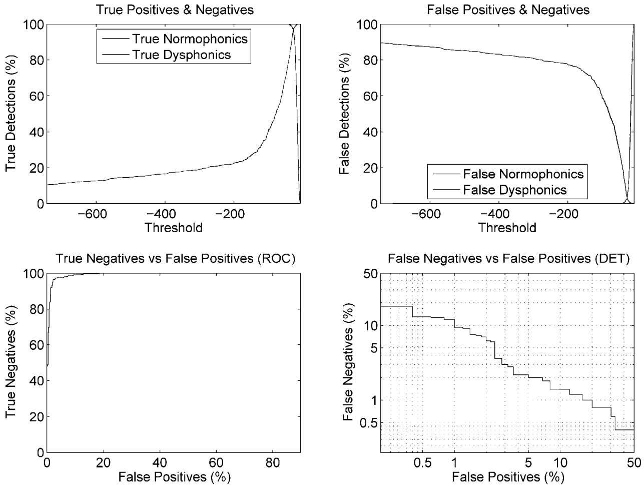


**Figure 8.** Validation results for male normophonic and dysphonic sets. Top left: True detections as a function of nonlinear threshold. Top right: Tippett plots (complementary distribution). Bottom left: Resulting ROC curve. Bottom right: Equivalent DET plot.

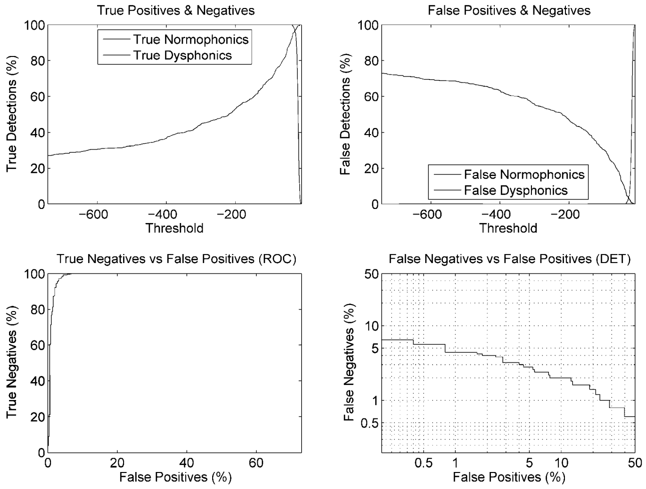


**Figure 9.** Validation results for female normophonic and dysphonic sets. Top left: True detections as a function of nonlinear threshold. Top right: Tippett plots (complementary distribution). Bottom left: Resulting ROC curve. Bottom right: Equivalent DET plot.

## 6. PRODUCTION AND DISTRIBUTION

The basic and applied research carried out to develop the algorithmics behind the G®E technology was initiated during the late 90's and took a decade to mature into some early applications mainly used with educational purposes in graduate and PhD programs. It was around 2006 where the first GUI's were developed to allow the use of the technology by researchers others than the developers in the medical and forensic fields. The cooperation with Hospital Universitario Gregorio Marañón, the Department of Criminalistics of Police Forces, and the Master in Forensic Sciences run by the Superior Council of Scientific Research (CSIC) opened the possibility to test these early GUI's with end-users. This resulted in the possibility of maturing and validating the basic G®E technology. But its exploitation required a different framework out of the scope of basic research. It was in this framework when the research group was motivated to present the G®E technology tied to a business plan to the VII Contest launched from Universidad Politécnica de Madrid in early 2010 to create Start-Up Companies with strong technological profile. The idea received the strong support of the Jury being granted the first prize among other 260 proposals [6]. In this way a new company under the name of BioMet®Soft was constituted at the end of 2011. The patent-protected technology [7, 8] transferred to the new company allowed the production of BioMet®Tools formally initiated in early 2012. The distribution is scheduled for the second half of 2012 under three different modalities: evaluation, cost-free (EVAL); academic, small-fee (ACAD) and professional, fully supported and maintained (PROF) [9].

## 7. CONCLUSIONS

Through the present paper a tool for the extraction of semantic information from the glottal source obtained from phonation has been introduced under the name of Glottal®Engine. Based on this technology a set of applications for the Detection of Dysphonic Voice, Forensic Voice Analysis or Education of the Singing Voice are being produced by BioMet®Soft, a start-up company created by Universidad Politécnica de Madrid to exploit the technology under Spanish and European Patent. It is expected that this experience will constitute a successful model to promote high-tech SME's devoted to Research-driven Innovation. The validation of the application BioMet®Phon for laryngological and speech therapeutic purposes has been tested on specific study cases, one of which has been discussed to a certain extent. Data from the technology validation tests have also been presented and discussed, showing the capabilities of the technology.

## 8. ACKNOWLEDGEMENTS

This research has been possible with the aid of grants TEC2009-14123-C04-03 and TEC2012-38630-C04-04 from the Ministry of Economy and Competitiveness of Spain.

## 9. REFERENCES

1. Gómez, P., Fernández, R., Rodellar, V., Nieto, V., Álvarez, A., Mazaira, L. M., Martínez, R, Godino, J. I.: Glottal Source Biometrical Signature for Voice Pathology Detection, *Speech Comm.*, Vol. 51, pp. 759-781 (2009).
2. Titze, I. R.: *Principles of Voice Production*, Prentice-Hall, Englewood Cliffs, NJ (1994).
3. G. Fant and J. Liljencrants: A four parameter model of the glottal flow, *STL-QPSR*, Vol. 26, No. 4, pp. 1-13 (1985).
4. Baken, R. J., Orlikoff, R. F.: Clinical Measurement of Speech and Voice, Singular Pub. Group, San Diego, CA (2000).
5. Martin, A., Doddington, G., Kamm, T., Ordowski, M., and Przybocki, M.: The DET curve in assessment of detection task performance. *Proc. Eurospeech 1997*, Rhodes, pp. 1895–1898 (1997).
6. Gómez, P., Fernández, R., Rodellar, V., Glottex®: Innovation in Biomedical Engineering for Life-Quality, *1st Int. Conf. On Quality and Innovation in Engineering and Management*, pp. 289-292, Cluj-Napoca, Romania (2011).
7. Patent No. P201131069 granted into Spain
8. European Patent No. PCT_ES2012_000137 (Priority P201131069)
9. http:\\www.glottex.com