Augmenting Sonic Perception with Synaesthesia in terms of Neurocognitive and Physiological Factors

D. Politis[*], V. Aleksić[†], G. Kyriafinis[**], A. Klapanaras[*]
{[*]Dept. of Informatics || [**]AHEPA University Hospital}, Aristotle University of Thessaloniki, Greece
[†]University of Kragujevac, Faculty of Technical Sciences, Čačak, Serbia

dpolitis@csd.auth.gr, veljko.aleksic@ftn.kg.ac.rs, , orlci@med.auth.gr

***Abstract*—Conjugate to Speech Science are Physiology, Anatomy, Linguistics, Physics, Computer Science, Psychology and so on, that provide tools and methodologies which can decipher the key elements of phonological activity in depth. Additionally, sonic perception provides in synaesthetic terms an acoustic augmentation due to the complex activity of tuned voicing with articulation-related physiological and neuronal dipoles. All these features affect the oral and aural communication channel and enhance the understanding of human communication in terms of Disordered Voicing, Vocal Affections, Music Experience and Human Attitudes.**

***Keywords*—Phonation, Singing, Oral and Aural Communication, Neurophysiology, Anatomy, Synaesthesia.**

I. INTRODUCTION

Human Communication relies heavily on how we perceive the acoustic signal. In the multimedia world we live in, in sensory terms, 89% of input is ascribed to vision, 10% to hearing, and 1% for all the other senses. However, in the practice of everyday communication, people utilize hearing, while assigned to some other tasks, at a rate significantly higher than the 10% that is credited to sonic stimuli. To give some examples, a listener receives a multitude of aural signals that trigger the sensory mechanisms for hearing when talking to his phone, when listening to music in his car or when harking at the discussions of his professional environment. Although he cannot see the sources of phonation, he may well imagine how the person singing or talking may look like in anatomical and neurophysiological terms.

Literally, synaesthetic perception promotes a union of senses: It is a neurological oddity, in which a sensation of one of the senses triggers involuntarily and automatically a parallel sensory or cognitive pathway [1]. It is a focal point of recent research in acoustics, relating chromatic reconstruction with musical impression.

Even further, as far as sensory augmentation and substitution are concerned, the last few years scientists have been arguing that humans work with a lot more than five senses, that they have some tenths of emotions, and that cerebral activity continuously shapes impressions of the world we live in [2].

This research explores the major factors that enable this synaesthetic augmentation. Neurophysiological studies provide complementary intelligence about the way that acoustic evidence indicates how speech signal variability relates with multimodal, kinetic and predominantly sparkling bodily sensations [3][4].

Music seems to be the most prominent triggering mechanism in terms of form, harmony and emotional arousing. It seems to offer a global substrate for the interconnection of oral and aural communication. Nevertheless, it is evident that speech communication plays a significant role as well, when the multitude of the so many spoken languages (some 7,200 worldwide) is taken into account [4].

Given that this research is very new, its line of investigation needs solid foundations. For this reason, it relies heavily on well-attested findings from the diachrony of music research. For instance, in 1993 James [5] had stated: "Music contains in its essence a mystery: everyone agrees that it communicates, but how?"

Music cognition therefore has emerged the last 40 years as an interdisciplinary field that brings together scientific fields like cognitive psychology, artificial intelligence, neuroscience and linguistics [6]. As the focus of popular music shifts from instrumental to vocal renditions, cognitive processes may be described more easily in neurophysiological terms, although contemporary music is relentlessly genre-crossing, combining elements of free jazz, rock, pop, dance music, ballads, etc. Under this prism, we may understand the purpose of music as an interaction with an intellectual system that aims to "engage, inspire, and enthrall a group of musicians into doing music that they are excited about, so that that excitement is passed on to the audience [7]."

Therefore, although in neurophysiological terms, speech communication is the medium for delivering the message, when music is involved, our intellect forms more complex dialects, as they have been discovered thus far on studies of the Brain Computer Interface performance of post-deaf subjects, rehabilitated with Cochlear Implants [8]. So, apart from the lexical semantics of human communication, the "new music" revolution, as it is abundantly received via music for films, radio-emissions and televised live concerts, permeates many other sensations to the everyday sonic synaesthetic potential of the human oral-aural communication channel [9].

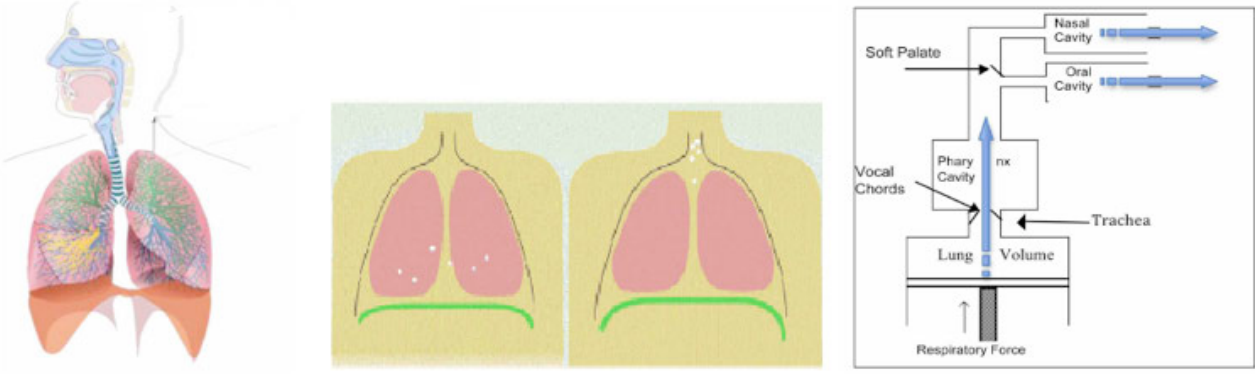

Figure 1. Modeling human phonation

As a result, when listening to a singer, listeners overwhelmingly embrace his overall *musical ability*. Via this they prolong their auditory imagery, sometimes with kinesthetic information [10]. In a reciprocal manner, from what listeners sense and feel, they may predict the state of euphoria in which the singer performs at the other end of oral-aural communication channel.

Within this context, some ten semi-professional singers were tested, singing contemporary tunes, most in English and some in Greek. Their evaluators try to decipher how the bodily, neurophysiological and social influences affect the auditory stimuli they receive, without having any other information on their "collective" build-up. Relying solely on the sonic impression, with "reverse-engineering" thinking, they attempt to reconstruct their actual world parameters.

But, as far as synaesthesia is concerned in singing, three significant precautions should be taken into account:

- Humans have their greater capacity for receiving physical feelings and perceptions when they are born. Their sensitivity decreases with age. After middle age their sensual awareness decreases considerably. Although people learn easier and in depth when they are young, singers and musicians deliver their best performances when they are mature and onwards.
- Singers, and generally speaking performers, have cultivated within the context of their special education the ability to replicate the physiological effects of many voices. More than usual people, they can give to their vocalizations a wide spectrum of sound characteristics and play many "roles" while performing. Typically, their everyday talking voice, which is closer to their neurophysiological build-up, does not coincide with their on-stage transformations. However, singing reveals the person's full range neurophysiological potential, and for this reason it is taken into account.
- For deciphering the neurophysiological characteristics of a singer, the compass of a wide range for mental and physical attributes is encountered, not merely a rhythmic or spectral analysis of the received audio signals. This research attempts to categorize and grade them.

II. Research Questions: How the Anatomy of the Sound Sources Shapes Phonation

Vocal sounds are in the broader sense all the sounds produced by a person's larynx and uttered through the mouth. They may be speech, singing voices, whispers, laughter, snorting or grunting sounds, etc.

Usually phonation serves as the basis for singing or communicating with each other. No matter how these sounds are produced and what communication purposes they may serve, they are categorized to:

I. Voiced sounds that are produced in a person's larynx with a stream of air coming from the lungs and resonating the vocal cords. This stream continues to be modulated through its passage from the pharynx, mouth, and nasal cavity, resulting to an utterance in the form of speech or song.

II. Unvoiced sounds. Singing or utterance would be incomplete if unvoiced sounds where not produced. They do not come as a result of a vibration from the vocal cords, but as partial obstruction of the airflow during articulation.

In terms of physiology, phonation is a complex process that involves many organs of the human body [4][10]. Some kinetic displacements, like those of the lungs and the chest are also affecting the final output. As a result, the major motor functions of phonation at this level are performed by muscular movements of the larynx, pharynx, vocal folds, mouth, tongue, jaws, nose and the respiratory system in general. Indeed, the muscles' motor action affects the vocal box position and tension to appropriate levels, so to adjust efficient respiratory force that maintains greater air pressure bellow the vocal chords than above them. This combined action of muscular settings, which unfortunately deteriorates with ageing, shapes the personal characteristics of voicing: loudness, pitch, roughness, smoothness, style, attractiveness, nerviness, elasticity, etc.

Schematically, these parameters are set up in Figure 1.

Research in this field dates back from the days of Sundberg [10], the KHT research group for Speech, Music and Hearing [11], and more recently Raphael et al. [12]. The most recent perspective can be found in [4], mingling the medical point of view with IT processing.

When music is encountered in vocal communication, the constraints of the linguistic message are somehow

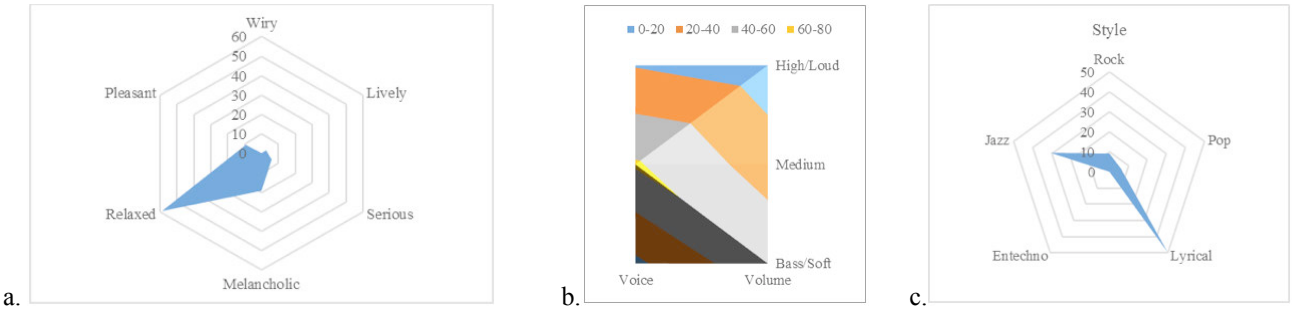

Figure 2. Singer V's (female) perceived characteristics

lifted, and listeners augment their synaesthetic awareness more or less at the same extent on trans-national basis.

Nonetheless, it is the first time that a synaesthetic association is attempted between the sounds we hear and the neurophysiological sketch-up of the source, as far as bodily construction is involved.

III. Methodology and Results

As previously stated, some 10 semi-professional singers were used, both male and female, singing various songs. Their mother tongue is Greek. Six very characteristic recordings were chosen, some in English and some in Greek.

The evaluators tried to figure out the anatomical, neurophysiologic and temperaments build up of the singers. They were even asked to describe how they imagine the upper part facial characteristics that depend on the nasal cavity with the nostrils. Indeed, this part plays a significant role in giving personified "images" of speech perceptual characteristics based on the articulatory and phonological characteristics [4].

The survey was conducted in January 2017. The evaluators were a pool of 32 students aged 20-25. Students participated anonymously, voluntarily and individually. They submitted electronically their anonymously processed reports.

The role of their tutor was to give initial instructions and ensure testing conditions. Testing was administered via the Aristotle University of Thessaloniki IT Center services.

The students that participated in the survey augmented their music perception by participating in a serious game for musical intelligence. They expanded their personal experience from singing and listening to songs with deduced reasoning and created a virtual profile of the singer. Al the test-bed songs were recorded in a studio. They range from pop style ballads up to some kind of hard rock renditions. This paper is the first step in a research that is bound to reveal a wealth of new bodily and psychokinetic features that affect the singing potential.

IV. Results and Discussion

In order to improve the transparency of the results, we will discuss them for each singer individually.

Singer V. (female)

The estimated age of singer named V. was 31.4 years (SD = 2.9), which is higher than actual (28). Students estimated that the singer was 167.7cm high (SD = 4.7), which is close to the actual height (169cm), but highly overestimated her weight (M = 73.3; SD = 7.1), which was actually 51kg. An independent samples t-test was conducted to examine whether there was a significant gender difference between students' estimation of the singer age, height and weight. The test revealed that there were no statistically significant differences between male and female students (p = .544, p = .450 and p = .560 respectively).

The t-test was also conducted to examine whether there was a significant gender difference between young evaluators in relation to their self-estimation of the perceived singer hair, eyes and mouth. A statistically significant difference was revealed between male and female students in perceived singer hair length (t = 1.08, df = 11.4, p = .304). Males (M = 1.95, SD = 0.22) perceived significantly longer hair length than did females (M = 1.80, SD = 0.42). In total, 53% of students correctly perceived brown hair color, 81% correctly perceived brown eye color. Most of the students imagined the singer having a medium size mouth (53%) with normal lips (59%).

When students were asked to determine V.'s mood,

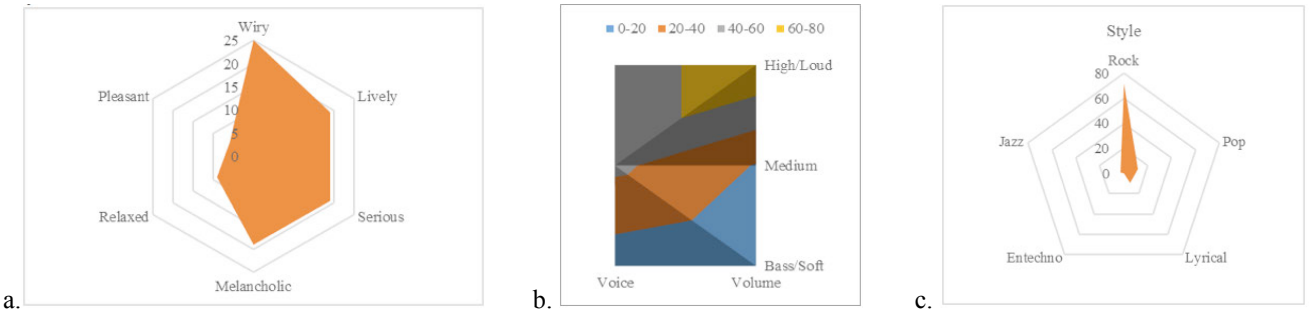

Figure 3. Singer G.'s (male) perceived characteristics

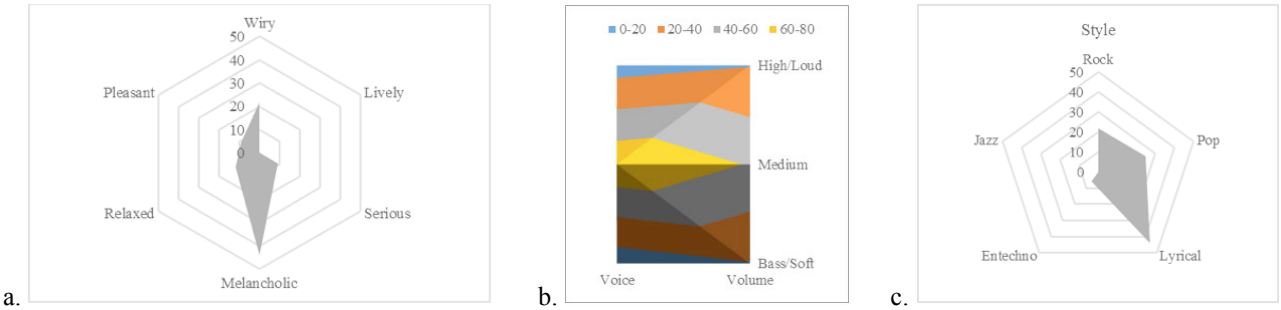

Figure 4. Singer S.'s (female) perceived characteristics

59% perceived correctly (calm/relaxed), but the rest were not far off as they described it as melancholic and pleasant. The results are presented on Figure 2, left.

V. has a calm, demanding, sweet-strenuous (loud) and tender voice. V.'s voice and volume perceived by students is presented on Figure 2, center and right.

A multiple linear regression was calculated to predict singers' voice profile (voice, style, volume) based on student gender, but no significant regression equation was found (p = .410, p = .319 and p = .642 respectfully).

Singer G. (male)

The estimated age of singer named G. was 27.5 years (SD = 3.2), which is higher than actual (21). Students estimated that G. was 172.2cm high (SD = 4.1), which underestimated the actual height (183cm). Similarly, they underestimated his weight (M = 77.5; SD = 4.1), which was actually 90kg. The t-test revealed that there were no statistically significant differences between male and female students estimation of the singer age, height and weight (p = .604, p = .855 and p = .216 respectfully).

Students mostly perceived the singer with long hair (90%), 87% imagined brown. A statistically significant difference was revealed between male and female students in perceived singer eye color (t = -2.50, df = 20, p = .021). Males (M = 2.52, SD = 0.87), perceived significantly more often green eye color than did females (M = 3.00, SD = 0.00). Each female student perceived G. with brown eyes, which was correct. Most of the students imagined that the singer has big size mouth (55%) with normal lips (78%).

When students were asked to determine G.'s mood, the results were colorful, as presented on Figure 3, left. This is due to the fact that G. has a very creative character.

G. has special voice quality (timbre). His voice and volume perceived by students is presented on Figure 3, center and right.

A multiple linear regression was calculated to predict the singers' voice profile (voice, style, volume) based on student gender, but no significant regression equation was found (p = .408, p = .399 and p = .317 respectfully).

Singer S. (female)

The estimated age of singer named S. was 29.5 years (SD = 2.9), which is much higher than actual (20). Students estimated that the singer was 169.1cm high (SD = 4.7), which is very close to the actual height (165cm), but extremely overestimated her weight (M = 78.0; SD = 5.1), which was actually 42kg. An independent samples t-test was conducted to examine whether there was a significant gender difference between students' estimation of the S. age, height and weight. The test revealed that there were no statistically significant differences between male and female students (p = .975, p = .988 and p = .854 respectfully).

Students mostly perceived the singer with the long hair (91%), but only 34% correctly imagined brown (more did black). Most students perceived the singer with brown eyes (91%). Most of the students imagined the singer having medium size mouth with normal lips (53%).

When students were asked to determine S.'s mood, 44% perceived incorrectly as melancholic. S. is passionate, but students did not perceive it. The results are presented on Figure 4, left.

S. has a passionate and sweet voice. S.'s voice and volume perceived by students is presented on Figure 4, center and right.

A multiple linear regression was calculated to predict S.'s voice profile (voice and volume) based on student gender, but no significant regression equation was found (p = .186 and p = .230 respectfully). However, when the same linear regression was calculated to predict S.'s voice style based on student gender, a significant

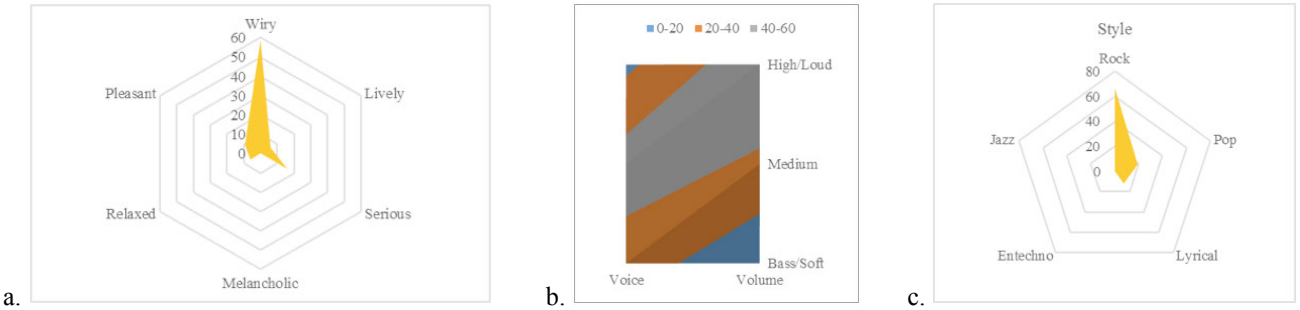

Figure 5. Singer A. 's (male) perceived characteristics

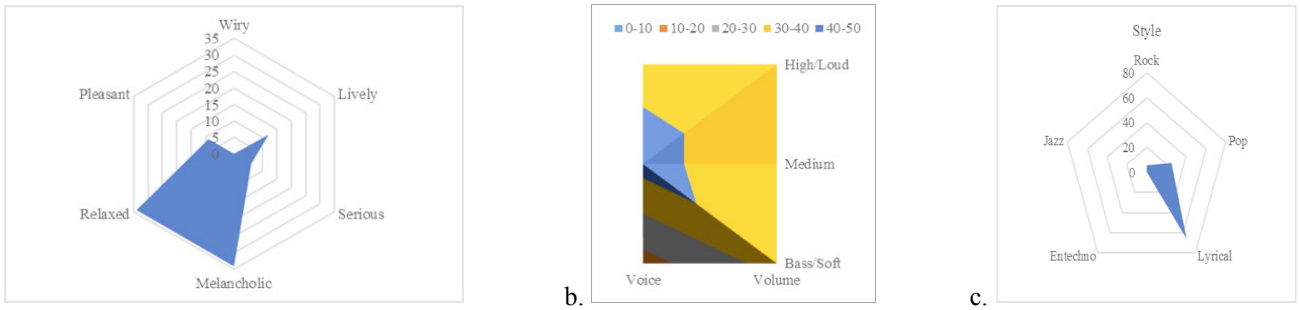

Figure 6. Singer T.'s (female) perceived characteristics

regression equation was found [$F_{(1, 29)} = 8.5$; $p = .007$; $R^2 = .228$]. The results show that student gender [$\beta = -1.1$; $t_{(29)} = -2.9$; $p = .007$] significantly predicted perceived singing style.

Singer A. (male)

The estimated age of singer named A. was 29.7 years ($SD = 2.6$), which is bit lower than actual (34). Students estimated that the singer was 174.6cm high ($SD = 4.6$) which is the actual height (175cm), but underestimated his weight ($M = 79.0$; $SD = 5.5$) which was actually 85kg. An independent samples t-test was conducted to examine whether there was a significant gender difference between students estimation of the singer age, height and weight. The test revealed that there were no statistically significant differences between male and female students ($p = .714$, $p = .112$ and $p = .440$ respectfully).

Students mostly perceived A. with short hair (59%), 44% correctly imagined brown. The statistically significant difference was revealed between male and female students in perceiving singer eye color ($t = 0.90$, $df = 11.8$, p = .387). Males ($M = 2.86$, $SD = 0.48$) correctly perceived significantly more often brown eye color than did females ($M = 2.60$, $SD = 0.84$). Students mostly imagined A. with big mouth and normal lips (53%).

When students were asked to determine A.'s mood, most incorrectly perceived him as wiry (59%), while 16% perceived him as serious. This maybe the consequence of the fact that A. has a very strong personality, so students wrongly perceived him as nervous. The results are presented on Figure 5, left.

A. has a raw and rough voice. His voice and volume perceived by students is presented on Figure 5, center and right.

A multiple linear regression was calculated to predict singer's voice profile (voice, style, volume) based on student gender, but no significant regression equation was found ($p = .738$, $p = .440$ and $p = .093$ respectfully).

Singer T. (female)

The estimated age of singer named T. was 28.2 years ($SD = 3.5$), which is almost actual (27). Students estimated that the singer was 169.5cm high ($SD = 4.1$), which overestimated the actual height (160cm), and highly overestimated her weight ($M = 78.3$; $SD = 6.5$), which was actually 50kg. An independent samples t-test was conducted to examine whether there was a significant gender difference between students' estimation of the singer age, height and weight.

The test revealed that there were no statistically significant differences between male and female students ($p = .898$, $p = .428$ and $p = .673$ respectfully).

Most students perceived T. with the long hair (78%) and 50% correctly imagined brown color. About 50% of the students incorrectly perceived T. with the blue/green eyes, as she has brown eyes. Most of the students imagined that singer having a medium size mouth (38%) with normal lips (69%).

When students were asked to determine T.'s mood, only 12% perceived correctly (lively). Most of the students incorrectly perceived her mood as relaxed/melancholic (69%). The results are presented on Figure 6, left.

T. has an assertive, energetic and dynamic personality, with thin and sinewy, shrilling voice. T. 's voice and volume perceived by students is presented on Figure 6, center and right.

A multiple linear regression was calculated to predict singer's voice and style based on student gender, but no significant regression equation was found ($p = .359$ and $p = .490$ respectfully). However, when the same linear regression was calculated to predict T.'s voice volume based on student gender, a significant regression equation was found [$F_{(1, 29)} = 7.2$; $p = .012$; $R^2 = .199$]. The results

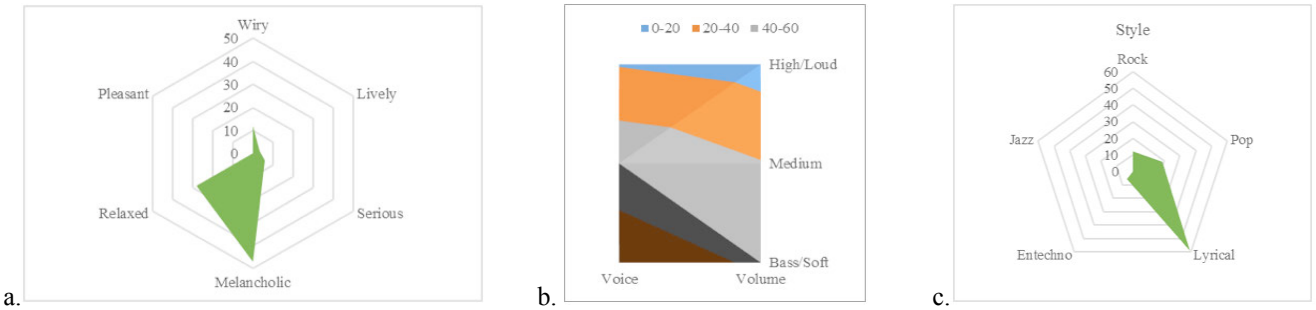

Figure 7. Singer Y.'s (female) perceived characteristics

show that student gender [β = 0.79; $t_{(29)}$ = 2.7; p = .012] significantly predicted perceived singing volume.

Singer ***Y. (female)***

The estimated age of singer named Y. was 28.9 years (*SD* = 3.3), which is excellent, as actually it is 28. Students estimated that the singer was 170.1cm high (*SD* = 4.8), which indeed is the actual height (170cm), and slightly overestimated her weight (*M* = 79.0; *SD* = 6.5), which actually is 75kg. An independent samples t-test was conducted to examine whether there was a significant gender difference between the students' estimation of the singer age, height and weight. The test revealed that there were no statistically significant differences between male and female students (p = .237, p = .091 and p = .577 respectfully).

The statistically significant difference was revealed between male and female students in perceived singer hair length (t = 1.13, df = 14.4, p = .277). Males (*M* = 1.81, *SD* = 0.40) correctly perceived significantly more often long hair than did females (*M* = 1.60, *SD* = 0.52). Only 37% of students correctly imagined brown color hair. Most of students correctly perceived the singer with brown eyes (84%). About half of the students imagined that singer having medium size mouth with normal lips.

When students were asked to determine Y.'s mood, only 19% perceived correctly (serious/wiry). Most of the students incorrectly imagined her as melancholic and relaxed, probably because she has a very sweet voice. The results are presented on Figure 7, left.

Y. has an assertive, energetic and dynamic personality, and the same time a sweet, bass voice. Y.' s voice and volume perceived by students is presented on Figure 7, center and right.

A multiple linear regression was calculated to predict singer' s voice profile (voice, style, volume) based on student gender, but no significant regression equation was found (p = .854, p = .929 and p = .229 respectfully).

V. Conclusion

With a rough estimate around 75% evaluators may decode the singer's voice and connect it with the actual neurophysiological characteristics of his/her bodily construction and personality. Some characteristics, like the color of the eyes, are rather irrelevant with the singing capacity, but the evaluators prove to be good gamblers, obviously extrapolating data from their ethnological habitat.

Generally speaking, some parameters may be strongly dependent on ethnological, linguistic and cultural biases. For instance in Serbia singers would have an increased percentage of blue eyes, while in Greece most of them have brown. Knowing the nationality of the singer strongly biases such a prediction.

In any case, the performer exhibits an exaggerated, augmented behavior and attempts to surpass the limits of his repertoire. Being well trained, he attempts to deliver a stable musical and artistic rendition. The performer tries to adopt himself to the singing role's prerequisites. Although some neurological factors, like mood and emotion, may be imitated, nonetheless the singing voice reveals a common denominator linked with bodily construction, personal, cultural and psychological ego (Figure 8).

These parameters can be more or less predicted.

References

[1] M. Bedny, A. Pascual-Leone, S. Dravida, and R. Saxe, "A sensitive period for language in the visual cortex: Distinct patterns of plasticity in congenitally versus late blind adults", *Brain Lang*., 2012, 122(3), pp. 162–170.

[2] S.J. Blakemore and S. Choudhury, "Development of the adolescent brain: implications for executive function and social cognition", *J Child Psychol Psychiatry*, 2006, 47(3-4), pp. 296-312.

[3] O. Bai, V. Rathi, P. Lin, D. Huang, H. Battapady, D. Y. Fei, L. Schneider, E. Houdayer, X. Chen and M. Hallett, "Prediction of human voluntary movement before it occurs", *Clinical Neurophysiology*, 2011, 122(2), pp. 364–372.

[4] D. Politis and M. Tsalighopoulos, "Oral and Aural Communication Interconnection: The Substrate for Global Musicality", in D. Politis, M. Tsalighopoulos, I. Iglezakis (Eds), *Digital Tools for Computer Music Production and Distribution*, IGI Global, PA:Hershey, USA, 2016, pp. 1-24.

[5] J. James, *The Music of the Spheres - Music, Science and the Natural Order of the Universe*, Grove Press, USA, 1993.

[6] D. Temperley, *The Cognition of Basic Musical Structures*, The MIT Press, MA:Boston, USA, 2001.

[7] J. Zorn, "The Game Pieces", in C. Cox and D. Warner (Eds), *Audio Culture - Readings in Modern Music*, The Continuum International Publishing Group, NY:New York, 2007.

[8] D. Politis, M. Tsalighopoulos, and G. Kyriafinis, "Dialectic & Reconstructive Musicality: Stressing the Brain-Computer Interface", in Proceedings of the *2014 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL2014)*, 13-14 November 2014, Thessaloniki, Greece.

[9] D. Deutsch, *The Psychology of Music*, .2nd Ed., Academic Press, USA, 1999.

[10] J. Sundberg, *The Science of the Singing Voice*, Northern Illinois University Press, USA, 1987.

[11] J. Iwarsson, M. Thomasson and J. Sundberg, "Lung Volume and Phonation: A Methodological Study", *Logopedics Phoniatrics Vocology*, 1996, 21(1), pp. 13-20.

[12] L. Raphael, G. Borden, and K. Harris, *Speech Science Primer - Physiology, Acoustics and Perception of Speech*, Lippincot, Williams & Wilkins, USA, 2007.

Figure 8. Singers V. and T. in their everyday communication (a. and c.) and in their artistic transformation (b. and d.)