

Semiotics of the non verbal vocal expression of emotions and research into the psychotherapy process: a pilot study

by Luca Campanelli*, Elisabetta Ibernì†, Diego Sarracino‡, Silvia Degni§, Rachele Mariani†

1. Introduction

«Linguistics can tell us the number of phonemes that make up the moneme /*tonight*/, what its distinctive features are, how they are linked in syntagma, which semantic rules say that this moneme can be considered a lexical form or *lexeme* referring to a meaning, how this meaning – being a *sememe* – can be further analysed into its semantic components, whether they are *semi* or not ... But it cannot explain how and why this “word” can be pronounced in different ways, with such intonations, pauses, rhythms or inflexions that it can mean, at different times, “tonight you’ll pay for this”, or “I can’t wait for tonight to come”, or yet again “you really think it will be for tonight?!”» (Eco & Volli, 1970, p.6).

Umberto Eco and Ugo Volli are a good introduction to the subject of this paper: the plurality of sense in verbal communication and the autonomous semantic properties of non verbal vocal communication. We shall discuss the possible semiotic relations between codes, and therefore their importance in the field of research into the psychotherapy process through the illustration of an empirical pilot contribution.

Although verbal language is one of the most powerful semiotic devices (Eco, 1975), there is no doubt that some non verbal units have a content of their own and that this content often cannot be translated into verbal units. Think of the precision with which a facial expression, a gesture or the tone of voice can convey a meaning; on the other hand, think of how difficult, if not impossible, it can be to translate into a non verbal code, of whatever type, the expression “the sun *still* rises”.

In the formalization made by Garroni (1973), given two sets of contents conveyable respectively by linguistic (L) and non linguistic devices (NL), there are intersecting portions where it is possible to translate contents from L to NL and viceversa, but there remain vast portions of L and NL that are untranslatable (Figure 1).

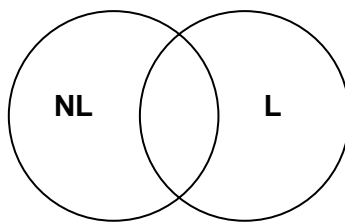


Figure 1. Diagram of the model proposed by Garroni (1973).

It seems that it was precisely this impossibility of translating semantic areas typical of non verbal codes into words that prompted Wittgenstein to overturn some of the convictions he expressed in the *Tractatus*; namely, in simple terms, to move from an abstract, mathematical idea of language where verbal categories are the only possible way of representing the world (Wittgenstein, 1922),

* Psychologist, University of Rome “La Sapienza”.

† Psychologist and psychotherapist, PhD student in dynamic and clinical psychology, University of Rome “La Sapienza”.

‡ PhD in dynamic and clinical psychology, research scholarship holder at the faculty of Psychology 1, University of Rome “La Sapienza”.

§ Psychologist, currently at the 2nd School of Specialization in Clinical Psychology, University of Rome “La Sapienza”.

towards a conception of language as an open, plural “game”, a set of expressions to study in their functional multiplicity and in their real manifestations (Wittgenstein, 1953). Now, this change of thinking seems to have been inspired by a conversation with Piero Sraffa. During a train journey, the Italian economist challenged Wittgenstein to translate the meaning of a Neapolitan gesture. This was a rapid movement of the back of the hand beneath the chin which actually expressed more than one meaning. It was soon clear to the Viennese philosopher that it was impossible not only to verbally translate the meaning of the gesture, but also to adequately describe its signification.

What we wish to underline here is that concrete communicative acts are always the outcome of different components, verbal and non verbal, arising in a specific cultural and relational context. The verbal and non verbal expressions are present in a complementary way and contribute to defining the sense of the utterances through their combinations. Seeking a semiotics of the paralinguistic code, and sustaining its autonomy, therefore does not mean contrasting it to other codes. It is in fact difficult to imagine a world without verbal language, where human beings communicate through dance, pointing to objects and uttering amorphous sounds. It is equally difficult to imagine human interaction composed of words alone.

However, we must point out, as we said earlier, that our interest is addressed not to the heterogeneous world of non verbal expressions, but exclusively, in this paper, to paralinguistic, i.e. a set of vocal and nonverbal signs such as loudness, tone, rhythm and voice timbre. In particular we focus on two types of relations involving paralinguistic:

- a) relations with the vocal verbal text of the communication;
- b) relations with the emotions, which we will examine in the next section.

As far as relations with the verbal are concerned, we can identify at least four types of functions performed by paralinguistic (Anolli & Ciceri, 1992, p.116):

- 1) *Hyper coding*. Non verbal vocal parameters can clarify and specify the meaning of verbal units through variations of loudness or pitch.
- 2) *oppositive* function. In this case the acoustic characteristics of the voice are in contrast with the meaning of the verbal units. Think for example of irony, where the meaning comes from the contrast between the verbal and non verbal message.
- 3) *Modification* of the meaning conveyed by the verbal code. This is not a matter of specifying or completing the verbal message, but of shifting the meaning to a greater or lesser degree. We can think for example of the same statement uttered in a tone of doubt or certainty.
- 4) *Intensification* or *deintensification*, therefore the attenuation or amplification of the meaning of the text, as in the case of a rebuke made harsher or milder by the loudness of the voice.

2. *The vocal expression of the emotions: brief overview of the research by categories*

Emotional communication is a central semantic area in human communication, since it plays a major role in regulating the individual's relations with other individuals and the environment. In the study of the behavioural expression of emotions, in-depth studies have been made of facial expression, proxemics, posture and paralinguistic signs. We will examine the latter.

The first research into the vocal expression of the emotions dates back to the early 1900s, with the attempt by some psychiatrists to improve the understanding and diagnosis of emotional disorders by using new methods of electroacoustic (Isserlin, 1925; Scripture, 1921). Systematic research programmes on the effects of emotions on the voice only began however in the 1960s, with psychologists' renewed interest in the expression of emotions (Izard, 1971; Tomkins, 1962) and with the introduction by phoneticians and engineers of increasingly precise and sophisticated technologies for the registration and acoustic analysis of the voice.

Research in the field of the non verbal vocal expression of the emotions can be subdivided into three areas:

- a) studies on the processes of vocal encoding of the emotions;
- b) studies on decoding;
- c) studies on the mechanisms of inferring emotions.

2.1 *Studies on the encoding phase*

Studies on the processes of vocal encoding of the expression of emotions have tried to identify the configurations of acoustic parameters of the voice characterizing different types, qualities or dimensions of emotions; in other words, whether specific emotions correspond to different configurations of acoustic parameters.

Three research paradigms can be identified in this area:

1) Studies on *natural non verbal vocal expression* (Cowie & Douglas-Cowie, 1996; Johannes, Petrovitsh Salnitski, Gunga, & Kirsch, 2000; Williams & Stevens, 1969, 1972). This research used materials produced spontaneously in natural conditions: radio or television reports by journalists containing emotional reactions; recordings in exceptional conditions between pilots and the control tower and so on. Williams & Stevens (1972), for instance, analysed the voice of a journalist in a live broadcast from Lakehurst (New Jersey) describing the approach of the Hindenburg airship, when it caught fire.

This research, which seems to be an ideal design from the point of view of ecological validity, is however not free of methodological and technical/practical problems. The voice samples for instance are often brief, usually involve very few subjects and the sound quality of the recordings is not always good. It is also difficult to establish in hindsight the emotion vocally expressed by the speaker, since different subjects may respond differently to the same stimulus situation. Lastly, the influence of the verbal content on the paralinguistic cannot be checked since the texts change with each expression of emotion collected.

2) Studies on *experimental induction* of emotions (Bachorowski & Owren, 1995; Karlsson et al., 1998; Markel, Bein, & Phillis, 1973; Scherer, Feldstein, Bond, & Rosenthal, 1985). Another strategy that has been used to study the effects of emotions on vocal expression is that of inducing emotional states. In these cases use was made of stimuli consistent with the emotion to be expressed. Most of the research in this case adopted indirect methods like the induction of stress through particularly difficult tasks and the showing of films or images with a strong emotional impact. In the work of Markel, Bein and Phillis (1973), for instance, the subjects were asked to describe the plates of the Thematic Apperception Test aloud; an assessment was then made of the correlation between the degree of hostility emerging from the test and the acoustic characteristics of the narrations.

This approach is favoured by experimental psychologists for the degree of control over the variables that it allows, but there are disadvantages. Firstly the emotions induced are generally of a low intensity; furthermore they are not necessarily the same for all the subjects, since stimulus variables can arouse different reactions in subjects and different degrees of reaction.

3) Studies on the *simulation* of emotions (Banse & Scherer, 1996; Bezooijen, 1984; Cosmides, 1983; Davitz, 1964; Klasmeyer, 1999). In this research situation, the subjects, sometimes professional actors, are asked to read a sentence of generally neutral content, or a series of meaningless numbers or letters, each time expressing a different emotion.

Simulated vocal expressions of emotion are certainly more intense and clearer than those that are induced or produced naturally. It is however equally obvious that such expressions are more stereotyped. It is possible, in other words, that the actor-subjects emphasise obvious aspects, prototypical in their culture, while neglecting and underestimating other characteristics present in the natural expression of emotions.

Williams and Stevens (1972) compared a radio broadcast of the Hindenburg disaster with a simulated narration of the same scene. The acoustic analysis of the two recordings showed the same paralinguistic profile but with a clear accentuation of the acoustic parameters in the simulated version. This is confirmed by the results of other research works which reveal systematic differences between natural and simulated expressions of emotions with greater redundancy and less ambiguity in the latter (Allen & Atkinson, 1981; Motley & Camden, 1988).

In any case, we can argue like Scherer (1986; Banse & Scherer, 1996) that the expression of emotions is always "acted" to a certain extent, i.e. that there is no expression of emotion outside the socio-cultural context in which we are immersed with all the constraints and rules that it imposes. Moreover, when vocal representations of simulated emotions are consistently recognised by listener-judges, it can be argued that they reflect at least in part the expression of emotions in natural conditions.

2.2 Studies on the decoding phase

Research into the decoding phase has studied the processes of recognition of emotions based solely on the paralinguistic signals in communication (Bezooijen, 1984; Brown, 1980; Burns & Beier, 1973; Scherer, 1989; Scherer, Banse, Wallbott, & Goldbeck, 1991; Williams & Stevens, 1981).

The typical research design consists of asking actors to act out a certain number of different emotions through the production of language utterances of standard verbal content or completely meaningless content. A group of listener-judges has the task of recognising and labelling the emotions, choosing from a given series of alternatives.

Different reports on the decoding phase have shown an average recognition rate of about 60% (Bezooijen, 1984; Scherer et al., 1991).

A serious limit of this type of study is their tendency to study the discriminative capacities of the subjects (choosing between a given series of alternatives) rather than the accuracy of the recognition. This bias can be overcome by correcting the accuracy coefficients for the answers guessed at random; there is however no single accepted way of doing this (Banse & Scherer, 1996).

Recently, Scherer, Banse, and Wallbott (2001) reported the results of a cross-cultural study which revealed an average accuracy in recognising emotions of 66%, with slightly lower percentages for progressively greater cultural diversity among the subjects. This can be interpreted as supporting the hypothesis of the existence of universal rules governing the inferring of emotions based on the acoustic properties of the voice.

2.3 Studies on inference

While decoding research examines the subjects' ability to recognise a speaker's emotional state, studies on inference are interested in the processes of inference underlying the recognition of emotions from the voice. In other words, they try to answer the question: on what acoustic parameters of the voice are inferences of specific emotions based?

The research designs of these studies envisage the systematic variation of the acoustic indicators of the utterances presented to the subjects and the analysis of its effect on the mechanisms of inferring emotions.

In this context three different research paradigms can be identified:

1) *Correlation* between acoustic parameters of the voice and listeners' opinions (Banse & Scherer, 1996; Bezooijen, 1984; Scherer, Koivumaki, & Rosenthal, 1972; Scherer, Rosenthal, & Koivumaki, 1972). To identify the vocal characteristics that probably determine inferred judgement, measurements were made of the correlations between acoustic features of the vocal expression of emotions and the opinions of listener-judges about the speaker's emotion. Banse and Scherer (1996) identified 9-10 acoustic parameters (average F0, F0 sd, utterance speed, and others) which should explain most of the variance of the opinions given by the listener-judges.

One of the main limits of this method is to be found in its reliance on the type of vocal utterances used, as well as the fact that the results are strongly influenced by the quality of the recording and by the precision of the acoustic analysis.

2) *Disguising* of acoustic signals (Brown, 1980; Friend & Farrar, 1994; Scherer, Feldstein, Bond, & Rosenthal, 1985; Scherer, Ladd, & Silverman, 1984). In these studies the acoustic parameters of the voice are disguised, distorted or removed from the utterance, then the effects on the mechanism by which the listener-judges infer emotions from the modified material are explored. Numerous techniques are adopted in the literature: filtering, cutting and random remounting, inversion. Each of these techniques alters or removes certain acoustic signals leaving others unchanged. In any case the verbal content is no longer comprehensible.

The use of these techniques has been criticised by some researchers (Ochai & Fukumura, 1967) since, in their opinion, the total lack of intelligibility, being an exceptional experience, supposedly inhibits the process of recognition.

3) Methods of *synthesis* (Breitenstein, Van Lancker, & Daum, 2001; Cahn, 1990; Murray, Arnott, & Rohwer, 1996; Scherer & Oshinsky, 1977). Modern methods of synthesis in the field of language technology have provided researchers in the area of the vocal expression of emotion with new tools for the study of the effects of vocal signal manipulation on listeners' assessments. The results of these studies confirm the relative force of specific acoustic parameters in the attribution of emotion by the listener-judges.

2.4 Discussion

The results of the research reported confirm the hypothesis of the semantic autonomy of paralinguistic features, i.e. a relation between emotional meaning and acoustic parameters of the voice. As we have already partly underlined, each research paradigm described has advantages and drawbacks. It will be interesting in the future, when we have more data available, to make systematic controls on the convergence between the different approaches (Johnstone & Scherer, 2000).

By comparing the research related to the three different areas considered (encoding, decoding, and inference) one finds, in spite of the heterogeneity of methods and research paradigms adopted, that the results obtained are substantially homogeneous, evidence of the common nature of the encoding rules of the expression of emotions. The most coherent data emerging concern the levels of emotional arousal which seem to be indicated reliably by some of the acoustic parameters of the voice examined, such as fundamental frequency and loudness. To be able to describe the acoustic configurations characterizing the different emotions it will be necessary to consider a greater number of acoustic parameters.

Underlining the syntactic and semantic autonomy of the paralinguistic code from the verbal code does not however mean arguing for its independence. In the communication act the two codes are closely connected, first of all because they share the same vocal-uditive channel. The interdependence of the two codes also derives from the fact that paralinguistic features are expressed "over" the phonemes which in isolation or in succession form the words of a language. We must also remember linguistic functions, like accent, and pragmatic functions such as paraverbal features' regulation of turn-taking.

It is therefore no longer tenable to attempt to establish a hierarchy among the codes. The vocal non verbal is not secondary to the verbal, nor is it a mere emotional colouring of it. At the same time, nor can the non verbal be considered the best channel for manifesting emotions. The sense of the communication act in its socio-cultural and relational context is in fact the result of the reciprocal and complementary participation of all the codes, including the verbal and the paralinguistic, and of their semiotic relations.

From this point of view, the contrast between *semantic information* and *affective information* loses meaning, the former conveyed by the verbal content and the latter by "how" the utterance is made through paralinguistic features. We have in fact seen that the paralinguistic code is able to convey meanings in a stable, commonly accepted way, and that "how" the utterance is made is endowed with a meaning of its own.

3. Empirical contribution

Based on the data emerging and the comments made on non verbal communication of the emotions, we identified the conceptual coordinates for the design and interpretation of the pilot study presented here.

In view of the fact that some vocal features are constantly correlated to physiological arousal, we examined the function and occurrence of emotional arousal, as it emerges from the acoustic analysis of the voice, in the psychotherapy process. We also matched the acoustic parameters of the voice that emerged with the style and content features of the verbal language and with other aspects of the therapeutic process.

The theoretical framework of reference comes from *Multiple Code Theory* by Wilma Bucci (1997a) and from *Therapeutic cycle models* (Bucci, 1997a, 1997b; Karasu, 1986; Mergenthaler, 1996; Mergenthaler & Bucci, 1999).

3.1 Materials and methods

We studied the psychotherapy process in two sessions (V and VI) of a young woman being treated with psychodynamic psychotherapy.

The acoustic analysis of the voice (Campanelli et al., 2006) was applied to the digital recordings of the sessions. The acoustic parameters considered are those indicated in the literature as being the best indicators of emotional arousal: Average F0, Average loudness, Time (ratio between speaking and silence).

The other factors used for the study of the therapeutic process are:

- *Referential activity* (RA) (Bucci, Kabasakalian-McKay, & RA Research Groups, 1992). This measures the referential connections between verbal and non verbal representations. The method of empirical evaluation is based on the hypothesis that a person in contact with his/her own emotional experience will be able to communicate it so as to evoke in the listener or reader a corresponding wealth of images and sensations.
- *DMRS* (Perry, Kardos, & Pagano, 1993). This method defines 28 defence mechanisms distributed in 7 levels of maturity organized hierarchically, on a scale from 1 to 7 (from “acting out defences” to “mature defences”).
- *Therapist intervention scale*, on a continuum from “expressive” interventions to “supportive” interventions, following the instructions provided by Gabbard (1990). We identified three ranges of intervention: supportive (confirming, advice and praise, empathic convalidation), intermediate (encouragement to elaborate) and expressive (interpretation, confrontation, clarification).
- *IVAT-1*, a tool elaborated by Colli and Lingardi (2001) aimed at exploring the therapeutic alliance in terms of “ruptures” of the alliance on the part of the patient and of “reparations” made by the therapist.
- *CCRT-LU*, a version of the CCRT (Luborsky & Crits-Christoph, 1990) developed by Albani, et al. (2002), which allows us to obtain a schematic description of the evolution of the patient’s relational issues. The wishes (W) and responses (R) of the patient (S) or of others (O) towards the patient herself (S) are coded (for instance, the code WOS indicates the patient’s wish that others do something for her).
- Finally we compared these indicators with the qualitative description of the case.

3.2 Presentation of the case-study

Valentina is a 24-year-old Philosophy student who was sent to the University Clinical Centre by a friend. She has been living in the town for four years, her parents live about 500km away, she has an elder brother. The reason for the request concerns a persistent panic attack disorder which began a year after she started to live away from home, despite pharmacological therapy.

The clinical interviews analysed with the protocol OPD, Operationalized Psychodynamic Diagnosis (OPD Group, 2002), reveal a low capacity for psychological insight regarding her own mental states and the predominance of psychic and anxious-depressive symptoms. Valentina’s relationships seem to be centred around her attempt to place herself at the centre of attention, asking for help and care and feeling permanently misunderstood and neglected. Others, including the therapist, tend to notice Valentina’s constant tendency to criticise, disparage and reject others (as well as herself), which leads them to move away or to allow her autonomy. The main conflicts at the time of the consultation concern, firstly, the care that Valentina passively requests while claiming to be self-sufficient. The second conflict concerns self-esteem: Valentina seems to swing between the tendency to disparage herself and idealise others and, in contrast, to feel that she is special, endowed with above average qualities and surrounded by mediocrities. Oedipal conflict is found to be present but not at a sufficient level to be assessed. The degree of integration of the psychic structure is average overall. In particular, the integrated perception of self and the object, the well integrated communicative capacity leads the clinician to see psychotherapy with a psychodynamic orientation as suitable. The most disorganized areas concern emotional self-regulation and defences. Defence mechanisms are on the whole organized around a neurotic level, but show an intense use of narcissistic defences and denial when Valentina talks about herself.

SWAP 200 (Western, Shedler, & Lingardi, 2003) reveals a *histrionic style* personality (Q-T histrionic: 63.74; Q-T high functioning: 56.63; Q-T obsessive: 56.52; Q-T dysphoric: 52.23). The protocol of the Adult Attachment Interview (Gorge, Kaplan, & Main, 1985) confirms the psychodiagnostic information illustrated. During the narration of a loss, dating back to about ten years before, Valentina shows a decrease in metacognitive processes and a sudden disorganization of defences. The category attributed to the state of mind is that of Disorganized/Unresolved. On the whole, the ability to fluently narrate her own story and relational experiences during childhood enables a coherent, authentic image of Valentina's past to be constructed. The second attribution concerning state of mind about attachment is Secure sub-category F4, which suggests Valentina's tendency to be introspective and to imagine the mental states of herself and of others.

3.3 Results: session 5

Session 5 (Figure 2) shows a specific fluctuation of RA, well-known to clinicians and researchers, which can be interpreted in terms of the Multiple Code Theory (Bucci, 1997a). In the literature a correspondence has been found between sessions with high levels of integration, encoded by skilled clinicians, and a specific movement of the RA variable (Freedman, Lasky, & Hurvich, 2002). High level integration refers to sessions in which a process of consolidation is seen, along with a process of development and relatively stable exploration activity. It could reflect a set of functions, such as explicit transference, affective communication, a reflective function and a received interpretation, (Freedman, Lasky, & Hurvich, 2002). The session thus defined could also be described in terms of the concept of "good hour" (Kris, 1956) or Type 2 session (Green, 1999).

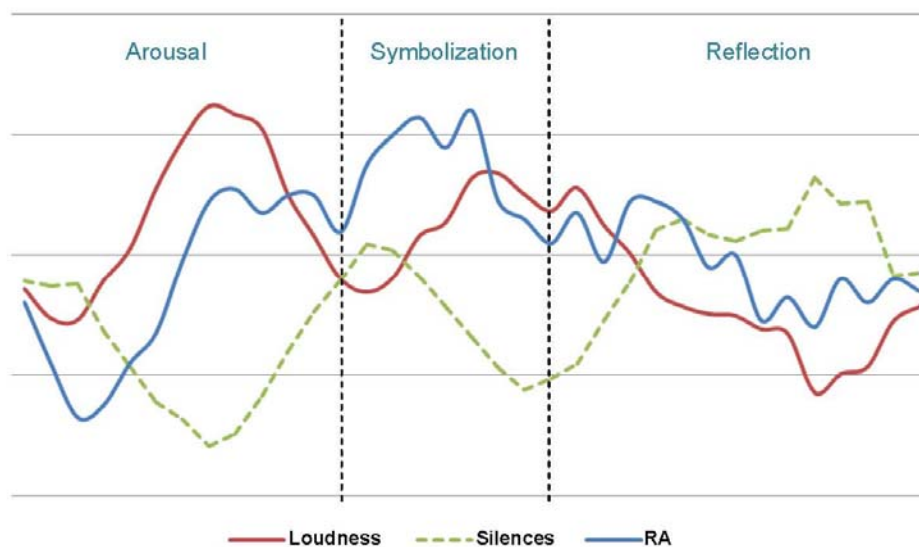


Figure 2. Session 5.

Bucci and Freedman (in press) have shown in these sessions that the RA tends to take on a typical bell shape: low values in the initial phase, progressively higher in the central stage and low again in the final stage of the session. According to the *Multiple Code Theory* this shape entails the arousal in the patient of an emotional pattern which is activated at the beginning of the session and which can be progressively narrated and re-elaborated. According to Bucci (1997a) the emotional pattern is encoded through sub-symbolic, symbolic non-verbal and symbolic verbal systems. The arousal of an emotional pattern always involves a subsymbolic arousal, which is more closely connected to visceral-somatic transformations. The activation of a pattern is therefore related to the lowering of the RA, where the visceral-somatic component is aroused, causing emotional arousal as yet detached from the direct narration. If one observes the progress of the RA in relation to the information deriving from the acoustic analysis and the analysis of defences, in the very first exchanges in the session one finds voice loudness and RA moving parallel and silence moving

inversely. This may indicate that in this phase the patient communicates in a discontinuous way, alternating between silence and words, due to the presence of more primitive defences. At the same time, the loudness of the voice diminishes, as if to indicate more an introspective dialogue, affectively connected to one's mental state, rather than a discourse with the interlocutor. The situation later changes, with greater emotional arousal, and the narration develops; moreover, the presence of primitive defences diminishes, allowing for more fluent communication. After the peak intensity, it can be seen that while the loudness of the voice falls, the RA continues to grow indicating a narration stage. The emotional pattern according to the Multiple Code Model is at this point aroused and linked to a memory or to a representation that is being recounted. The narration is slower, as can be seen by the increase in silences, as if the patient could relive the experience while narrating it. The end of the symbolization phase takes the form of a realignment of the RA and the loudness curve. This time both assume a falling movement, while the silences increase. This determines the entrance to the third phase in which silence prevails over communication and where the volume of the voice becomes intimate, indicating the deactivation of the emotional pattern and the re-elaboration of the same.

3.4 Results: session 6

In this section we present the results of the comparison between acoustic analysis, Referential Activity and other measures adopted for the study of the psychotherapies, concerning Valentina's sixth session. As can be seen from the graph (Figure 3) we divided the session into five sections, which correspond, from the point of view of the above-mentioned models of the therapeutic cycle, to five different stages of the session. For each section there are one or more "relational episodes", in terms of the criteria used in the CCRT encoding, designed to relate the content of the process to the formal approach.

1) 1st cycle: Arousal phase.

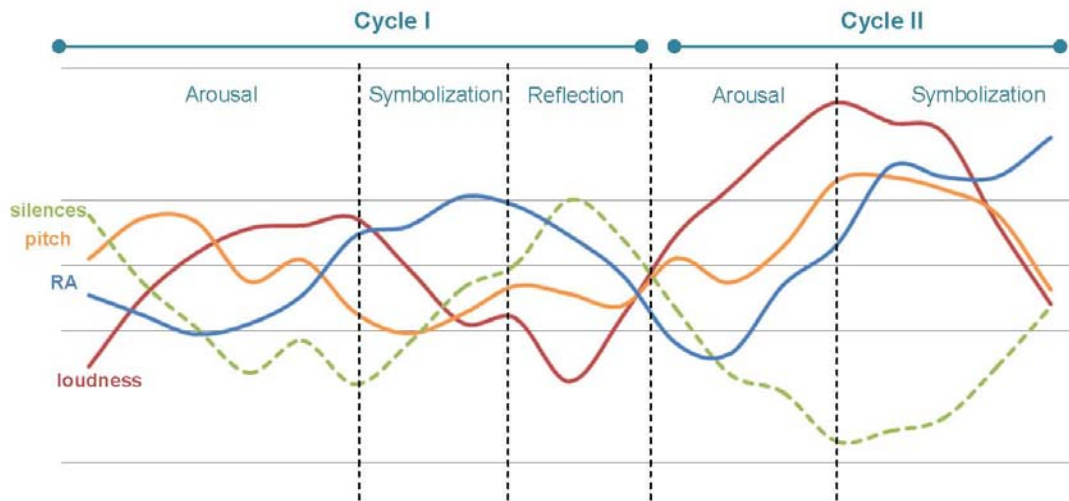
(Greetings)

In this first episode, which cannot be called a real relational episode according to the canons of the CCRT, the patient starts with an attempt to break up the therapeutic alliance, in which she tries to shift the focus from the *here and now* of the session; the therapist promptly makes an attempt at "reparation", bringing the discourse back to their present relationship. The CCRT categories, too, underline the patient's fear of "getting too close". In this first episode, the therapist's interventions are above all at an "intermediate" level: encouraging the patient to elaborate the material she brings to the session.

The low RA and the growing emotional arousal, associated with a very high silence rate, confirms the difficulty, hardly mentalized, of being in the relationship in this early stage of the session..

1st relational episode: herself

From this relational episode, the patient reports a series of problems about herself and in the later episodes, about her parents and "people" in general. When the patient talks about herself, she tends to use above all narcissistic type defences. The Central Relational Theme/Subject emerging from this and the later relational episodes is the desire, referring to herself and others, to be more "autonomous" (WSS D25) which is accompanied by a self-judging and self-punitive response by the self (RSS L23). In this episode, too, the therapist adapts his interventions to the narcissistic/neurotic level of the patient's defences, placing himself at an "intermediate" level in Gabbard's continuum.



	(greetings)	1. herself	2. parents	3. people	4. therapist	(clarificat.)	5. herself	(setting)	6. herself
CCRT	ROS near	WSS autonomous	WSO autonom.	WSO autonom.	WSS autonom.	RSS fear	RSS autonomous		WSS autonomous
LU	RSS fear	RSS withdrawal	RSO weak	ROS ignores	RSS criticises	RSS ignores	RSS admires		RSS criticises
DMRS		narcissistic	neurotic	neurotic acting-out	narcissistic neurotic	mature	narcissistic		narcissistic mature
IVAT-1	rupture/3 reparat./1			rupture/3 reparat./1					
Therap.	intermed.	intermediate	intermed.		intermed.	expressive	expressive	intermed.	supportive

Figure 3. Session 6.

2nd Relational episode: parents

The CCRT indicates that, also when reporting some recent interactions with her parents, the patient shows a desire to be “more autonomous” in this relationship (WSO D25), accompanied however by the feeling of being “too weak, too insignificant” (RSO G21). When talking about her parents, the patient’s defences shift to a neurotic level. The therapist’s interventions remain on an “intermediate” level, prompting her to elaborate the material further (for example, he asks her: “do you feel like talking about how this week at home went?”).

These first two relational episodes are characterized by a high emotional arousal and by RA scores that are still, though growing. The arousal phase is confirmed by very low percentages of silence and by a high speaking speed. The verbal content concerns her panic attacks (first relational episode) and the relationship with her parents (second relational episode), areas of experience that have not yet been adequately elaborated.

2) 1st cycle: Symbolization phase

Relational episode 3: “people”

The symbolization phase opens with this relational episode in which the patient refers to an impersonal object, considering her own situation in more general terms: “Whoever knows about this problem ... about feeling bad from let’s say an emotional point of view ... the solution the cure is to take your mind off it, not to think about it”. The IVAT signals a fresh attempt by the patient to bring about a break in the therapeutic alliance by shifting the focus from the here and now of the session, which is followed by the gradual attempt at reparation by the therapist (which will provoke the next episode centred on the patient-therapist relationship). The therapist tries to repair this breakdown by exploring the meaning of posing personal questions in the therapeutic context. The Central Relational Theme repeats and develops what emerged in the previous episodes, namely: “Others neglect me” (ROS I12), “I feel weak and insignificant” (RSS G21) and “I would like to be

stronger and more autonomous" (WSS D25). Talking about "others in general", it is significant to underline that the patient's defences are on the level of acting out, characterized above all by the use of passive aggression.

3) 1st cycle: Reflection phase

Relational episode 4: therapist

In this episode, characterized by a low level of emotional arousal, the patient begins a long discourse on the therapist. There emerge narcissistic and neurotic defences (disparagement and transfer), corresponding in the CCRT-LU to issues above all connected to herself: the patient feels that: "I could even find the answers, I mean, having a bit more self-confidence than I normally do, in the end I can give myself some of the answers". The therapist's intervention continues on the same "intermediate" level as the previous sections, namely, by tuning in to her defences, he encourages the patient to elaborate.

4) 2nd cycle: Arousal phase

(Clarifying intervention by the therapist)

At this point a new therapeutic cycle begins with an arousal phase, which will be followed by a further reflection phase (presumably due to a question of time, there is no reflection phase in this cycle). There is then an episode which, though it cannot be classed as a relational episode in the strict CCRT sense, is characterized by the therapist adopting a more expressive mode of intervention. In other words, he tries to reformulate what the patient reported in the previous episodes from a more coherent point of view, trying to give a meaning in the context of the therapeutic relationship. We can define this as the real "now moment" of the session (cfr. Stern, 2004). These crucial moments are, according to Stern, rich in affective and intersubjective experience. High emotional arousal, as it emerges from the analysis of the pitch and loudness of the voice, certainly seems to capture and confirm the interpretation of this phase of the session. We can see that the patient's defences become more mature (self-affirmation), although there persists a sense of fear related above all to the lack of control.

Relational episode 5: herself

In this episode there is the re-emergence of narcissistic issues that are evident both in the CCRT and in the DMRS. In the CCRT, the patient perceives herself as autonomous (RSS D25), and "admires" herself, demonstrating ambivalence towards what emerged in the previous episodes. Also the defences used are on this level (self-affirmation). The therapist's interventions, in the wake of the previous episode, remain on an expressive plane. The emotional arousal, high again, seems to confirm the centrality of the emerging narcissistic issues.

5) 2nd cycle: Symbolization phase

(Comments on the external setting)

After the emotional peak reached in the previous episode, there is a phase of symbolization involving the connection of emotional contents and specific episodes of experience. In this episode, which again cannot be considered a real relational episode according to the CCRT criteria, the attention is focused on some elements of the external setting, in particular on the question of "summer holidays". In any case, there is a return by the therapist to "intermediate", less expressive modalities of intervention, as if in preparation for the closure of the session.

Relational episode 6: herself

In this last episode the levels of emotional arousal are lowered and there is an increase in silence, as if to indicate a fading end to the session. Narcissistic aspects are still predominant, as is shown by the CCRT which takes up the question of self-criticism for not being more independent and for not having more self-confidence (WSS D25, RSS G12, RSS D26). The defence mechanisms are still at a narcissistic level but for the first time in an episode in which the patient talks about herself, mature defences also emerge. The therapist brings the session to an end with a supportive

intervention aimed at supporting the patient's progress and giving her confidence: "So perhaps there is time".

4. Conclusions

In the field of psychotherapy research, the study of the therapeutic process is based almost exclusively on transcriptions of the sessions alone. Only rarely have measurements been made of the non verbal behaviour of the patient, or even more rarely, of that of the psychotherapist. It is now clear however that the study of the verbal component alone, detached from the wealth of non verbal communication, reduces our chances of achieving an accurate understanding of the therapeutic relationship and the therapeutic process.

Based on this lacuna we have tried to describe the intersubjective emotional space by means of paralinguistic expression, adopting the acoustic indicators of the voice as the measurement of emotional arousal.

Even in this first, limited explorative application, the acoustic analysis of the voice has proved to be a particularly useful tool for assessing the importance for the patient of the subject of narration, and therefore the clinical centrality of the issues emerging and the way of treating them.

The analysis of the vocal expression of emotions can also help us to understand, along with other tools, to what degree emotional experience can be translated at the level of symbolic expression.

The main limit is to be found in the limited number of acoustic indicators of the voice examined: average F0, average loudness and silence-speaking ratio. These are the parameters that the literature indicates as most highly correlated to emotional arousal. The need remains for more sophisticated acoustic analyses that can take into account a greater number of acoustic parameters. Measurements of emotional arousal, as they emerge from the acoustic analysis of the voice, should also be systematically integrated and linked to the assessment of verbal language, and therefore to its style and content.

References

Albani, C., Pokorny, D., Blaser, G., Grüniger, S., König, S., Marschke, F., et al. (2002). Reformulation of CCRT categories: The CCRT-LU Category System. *Psychotherapy Research*, 12, 319-338.

Allen, V. L., & Atkinson M. L. (1981). Identification of spontaneous and deliberate behavior. *Journal of Nonverbal Behavior*, 5, 224-237.

Anolli, L., & Ciceri R. (1992). *La voce delle emozioni*. Roma: FrancoAngeli.

Bachorowski, J. A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychol. Sci.*, 6 (4), 219-224.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.*, 70 (3), 614-636.

Bezooijen, R. van (1984). *Characteristics and recognizability of vocal expression of emotion*. Dordrecht: Foris Publications.

Breitenstein, C., Van Lancker, D., & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cogn. Emot.*, 15 (1), 57-79.

Brown, B. L. (1980). The detection of emotion in vocal qualities. In H. Giles, W. P. Robinson, & P. M. Smith (Eds.), *Language: Social psychological perspectives. Selected proceedings of the first International Conference on Social Psychology and Language, held at the University of Bristol, England, July 1979* (pp. 237-245). Oxford: Pergamon.

Bucci, W. (1997a). *Psychoanalysis and cognitive science: A multiple code theory*. New York: Guilford Press.

- Bucci, W. (1997b). Patterns of discourse in "good" and troubled hours: A multiple code interpretation. *Journal of the American Psychoanalytic Association*, 45 (1), 155-187.
- Bucci, W., & Freedman, N. (in press). The interface of empirical research and analytic thought: the vision of Robert Wallerstein. In W. Bucci, & N. Freedman (Eds.), *The integration of Clinical work and Research Perspectives in Psychoanalysis: a tribute to the work of Robert Wallerstein*. New York: International Universities Press.
- Bucci, W., Kabasakalian-McKay, R., and the RA Research Groups (1992). *Scoring Referential Activity. Instructions for use with transcripts of spoken narrative texts*. Ulm: Ulmer Textbank.
- Burns, K. L., Beier E. G. (1973). Significance of Vocal and Visual Channels In the Decoding of Emotional Meaning. *Journal of Communication*, 23 (1), 118-130.
- Cahn, J. (1990). The generation of affect in synthesised speech. *J. Amer. Voice I/O Soc.*, 8, 1-19.
- Campanelli, L., Ibernì, E., Mariani, R., Sarracino, D., Degni, S., & De Coro, A. (2006, Settembre). *Fonetica acustica soprasedimentale: Un'applicazione esplorativa alla ricerca sul processo in psicoterapia per lo studio delle emozioni*. Presentato al V Congresso nazionale della sezione italiana della Society for Psychotherapy Research, Reggio Calabria.
- Colli, A., & Lingiardi, V. (2001). *Manuale IVAT: Indice di Valutazione dell'Alleanza Terapeutica*. Manoscritto non pubblicato.
- Cosmides, L. (1983). Invariances in the acoustic expression of emotion during speech. *J. Exp. Psychol.: Hum. Percept. Perform.*, 9, 864-881.
- Cowie, R., & Douglas-Cowie, E. (1996). Automatic statistical analysis of the signal and prosodic signs of emotion in speech. *Proc. ICSLP 1996 (1989-1992)*, Philadelphia.
- Davitz, J. R. (1964). *The Communication of Emotional Meaning*. New York: McGraw-Hill.
- Eco, U. (1975). *Trattato di semiotica generale*. Milano: Bompiani.
- Eco, U., & Volli, U. (1970). Introduzione all'edizione italiana. In T. A. Sebeok, A. S. Hayes, & M. C. Bateson. *Paralinguistica e cinesica*. Milano: Bompiani.
- Freedman, N., Lasky, R., & Hurvich, M. (2002, September). *Two pathways toward knowing psychoanalytic process*. Paper presented at IPA conference, Frankfurt.
- Friend, M., & Farrar, M. J. (1994). A comparison of contentmasking procedures for obtaining judgments of discrete affective states. *J. Acoust. Soc. Amer.*, 96 (3), 1283-1290.
- Gabbard, G. O. (1990). *Psychodynamic Psychiatry in Clinical Practice*. Washington, DC: American Psychiatric Press.
- Garroni, E. (1973). *Progetto di semiotica*. Bari: Laterza.
- George, C., Kaplan, N., & Main, M. (1985). *Adult Attachment Interview*, Unpublished manuscript. Berkeley: University of California.
- Green, A. (1999). The Fabric of Affect in the Psychoanalytic Discourse. *The New Library of Psychoanalysis*, 37.
- Gruppo di lavoro OPD (2002). *Diagnosi Psicodinamica Operazionalizzata*. Milano: Masson.
- Isserlin, M. (1925). Psychologisch-phonetische Untersuchungen. II. Mitteilung. *Z. Gesamte Neurol. Psychiatr*, 94, 437-448.
- Izard, C. E. (1971). *The Face of Emotion*. New York: Appleton-Century-Crofts.

- Johannes, B., Petrovitsh Salnitski, V., Gunga, H. C., & Kirsch, K. (2000). Voice stress monitoring in space: Possibilities and limits. *Aviat. Space Environ. Md.* 71, 9 (2), A58-A65.
- Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. Haviland (Eds.), *Handbook of emotion* (pp.220-235). New York: Guilford.
- Karasu, T. B. (1986). The specificity versus nonspecificity dilemma: Toward identifying therapeutic change agents. *American Journal of Psychiatry*, 143, 687-695.
- Karlsson, I., Bänziger, T., Dankovicova, J., Johnstone, T., Lindberg, J., Melin, J., et al. (1998). Speaker verification with elicited speaking styles in the VeriVox project. *Proc. RLA2C* (207-210), Avignon.
- Klasmeyer, G. (1999). Akustische Korrelate des stimmlich emotionalen Ausdrucks in der Lautsprache. In H.W. Wodarz, G. Heike, P. Janota, & M. Mangold (Eds.), *Forum Phonetikum* (Vol. 67). Frankfurt am Main: Hector.
- Kris, E. (1956). On Some Vicissitudes of Insight in Psycho-Analysis. *International Journal of Psycho-Analysis*, 37, 445-455.
- Luborsky, L., & Crits-Christoph P. (1990). *Understanding transference: The CCRT method*. New York: Basic Books.
- Markel, N. N., Bein, M. F., & Phillis, J. A. (1973). The relationship between words and tone of voice. *Language and Speech*, 16, 15-21.
- Mergenthaler, E. (1996). Emotions-abstraction patterns in varbatim protocols: A new way of describing psychotherapeutic processes. *Journal of Consulting and Clinical Psychology*, 64 (6), 1306-1315.
- Mergenthaler, E., & Bucci, W. (1999). Linking Verbal and non-verbal representations: Computer analysis of Referential Activity. *British Journal of Medical Psychology*, 72, 339-354.
- Motley, M. T., & Camden, C. T. (1988). Facial expression of emotion: A comparison of posed versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Speech Communication*, 52, 1-22.
- Murray, I. R., Arnott, J. L., & Rohwer, E. A. (1996). Emotional stress in synthetic speech: progress and future directions. *Speech Communication*, 20 (1-2), 85-91.
- Ochai, Y., & Fukumura, T. (1967). On the fundamental qualities of speech in communication. *J. of the Acoustical Society of America*, 29, 392-393.
- Perry, J. C., Kardos, M. E., & Pagano, C. J. (1993). The study of defenses in psychotherapy using the Defense Mechanism Rating Scale (DMRS). In U. Hentschel & W. Ehlers (Eds.), *The Concept of Defense Mechanisms in Contemporary Psychology: Theoretical, Research, and Clinical Perspectives* (122-132). New York: Springer.
- Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychol. Bull.*, 99 (2), 143-165.
- Scherer, K. R. (1989). Vocal correlates of emotion. In H. Wagner & A. Manstead (Eds.), *Handbook of Psychophysiology: Emotion and Social Behavior* (pp.165-197). London: Wiley.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *J. Cross-Cult. Psychol.*, 32 (1), 76-92.
- Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motiv. Emot.*, 15, 123-148.
- Scherer, K. R., Feldstein, S., Bond, R. N., & Rosenthal, R. (1985). Vocal cues to deception: A comparative channel approach. *J. Psycholinguist. Res.*, 14, 409-425.

- Scherer, K. R., Koivumaki, J., & Rosenthal, R. (1972). Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech. *J. Psycholinguist. Res.*, 1, 269-285.
- Scherer, K. R., Ladd, D. R., & Silverman, K. E. A. (1984). Vocal cues to speaker affect: Testing two models. *J. Acoust. Soc. Amer.*, 76, 1346-1356.
- Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motiv. Emot.*, 1, 331-346.
- Scherer, K.R., Rosenthal, R., & Koivumaki, J. (1972). Mediating interpersonal expectancies via vocal cues: Different speech intensity as a means of social influence. *Europ. J. Soc. Psychol.*, 2, 163-176.
- Scripture, E. W. (1921). A study of emotions by speech transcription. *Vox*, 31, 179-183.
- Stern, D. (2004). *The Present Moment*. New York: Norton.
- Tomkins, S. S. (1962). *Affect, Imagery, Consciousness: The Positive Affects* (Vol. 1). New York: Springer.
- Westen, D., Shedler, J., & Lingardi, V. (2003). *La valutazione della personalità con la SWAP-200*. Milano: Raffaello Cortina.
- Williams, C. E., & Stevens, K. N. (1969). On determining the emotional state of pilots during flight: An exploratory study. *Aerospace Md.*, 40, 1369-1372.
- Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *J. Acoust. Soc. Amer.*, 52, 1238-1250.
- Williams, C. E., & Stevens, K. N. (1981). Vocal correlates of emotional states. In J. K. Darby (Ed.), *Speech evaluation in psychiatry*. New York: Grune and Stratton.
- Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. London: Routledge & Kegan Paul.
- Wittgenstein, L. (1953). *Philosophische Untersuchungen*. Oxford: Blackwell.