

Rate of speech and emotional-cognitive regulation in the psychotherapeutic process: a pilot study

Marco Tonti,¹ Omar C.G. Gelo^{1,2}

¹Department of History, Society e Human Studies, University of Salento, Lecce, Italy; ²Faculty of Psychotherapy Sciences, Sigmund Freud University, Vienna, Austria

ABSTRACT

This study investigates the relationship between a client's rate of speech (ROS) and emotional-cognitive regulation during a psychotherapy session. The ROS was measured in words per second on the timed transcript of a single session of psychodynamic psychotherapy. Emotional-cognitive regulation was assessed using the therapeutic cycles model on emotional tone (ET), abstraction (AB), and their combination in emotion-abstraction patterns (EAPs). The results were mostly consistent with our hypotheses and showed that: i) the ROS negatively correlated with the conjoined ET and AB; and ii) the ROS in the connecting EAP (high ET and high AB) was significantly lower than in other EAPs. The results support the hypothesis that a significant reduction in the client's ROS may be a reliable marker of in-session change processes. Clinical implications and future developments are discussed.

Key words: Computer-assisted text analysis; Rate of speech; Therapeutic Cycles Model; Emotional-cognitive regulation; Psychotherapy process.

Introduction

Psychotherapy is a communicational enterprise realized through linguistic (semantics, syntax, and pragmatics), extralinguistic (gestures, gaze and eye contact, posture, spatial distance, *etc.*), and paralinguistic (prosody, including speech vocal cues as intonation, rhythm, speaking rate, loudness, pauses, voice quality,

etc.) channels; these three channels dynamically interact with each other during client-therapist communication.

There is increasing evidence in linguistics as well as social and clinical psychology (Jarmołowicz, Karpinski, Malisz, & Szczyszek, 2007; Keller & Tschacher, 2007; Wennerstrom, 2001) showing that the course of a dyadic interaction is shaped by all three channels (*e.g.*, Bernieri, 1989; Neumann & Strack, 2000; Núñez & Sweetser, 2006). Nonetheless, psychotherapy researchers have not paid equal attention to each of these three communicational channels. In fact, while linguistic and content analyses represent a well-established tradition in the field of psychotherapy research (Gelo, Salcuni, & Colli, 2012) and the analysis of extralinguistic communication is gaining increasing relevance (*e.g.*, Bänninger-Huber & Widmer, 1999; Ramseyer & Tschacher, 2011), the investigation of paralinguistic features in relation to the psychotherapeutic process is still not widespread.

This paper aims to provide empirical evidence on the relevance of prosody within the psychotherapeutic process. More specifically, we focus on a specific feature of paralinguistic communication, namely, the rate of speech (ROS), while the psychotherapeutic process is considered in reference to emotional-cognitive regulation. First, we briefly review the existing literature concerning the ROS and the psychotherapeutic process; then, we describe the Therapeutic Cycles Model (TCM) (Mergenthaler, 1996, 2008, 2015), which has been developed to analyze the therapeutic process in terms of emotional-cognitive regulation; finally, we outline the rationale of the present study.

Psychotherapeutic process and rate of speech

From a theoretical point of view, the clinical relevance of certain paralinguistic aspects of communication, such

Correspondence: Omar Carlo Gioacchino Gelo, Department of History, Society and Human Studies, University of Salento, via Stampacchia 45, 73100 Lecce, Italy.
Tel./Fax: +39.0832.294703.
E-mail: omar.gelo@unisalento.it

Acknowledgments: this work is based on the first author's Master's thesis in Computer Science, who is grateful for the support that he received from Prof. Erhard Mergenthaler and Prof. Marco Casonato. The authors are also grateful to psychotherapist Roberto De Ponte Conti for providing the recording of the studied session.

Citation: Tonti, M., & Gelo, O.C.G. (2016). Rate of speech and emotional-cognitive regulation in the psychotherapeutic process: a pilot study. *Research in Psychotherapy: Psychopathology, Process and Outcome*, 19(2), 102-112. doi: 10.4081/ripppo.2016.232

Received for publication: 5 July 2016.
Revision received: 27 September 2016.
Accepted for publication: 27 September 2016.

This work is licensed under a Creative Commons Attribution Non-Commercial 4.0 License (CC BY-NC 4.0).

©Copyright M. Tonti and O.C.G. Gelo, 2016
Licensee PAGEPress, Italy
Research in Psychotherapy: Psychopathology, Process and Outcome 2016; 19:102-112
doi:10.4081/ripppo.2016.232

as voice volume, pitch, and ROS within the psychotherapeutic process has been acknowledged for a long period of time (e.g., Bady, 1985; Wennerstrom, 2001). Rice and Kerr (1986) noted that *the vocal channel seems to be perhaps the most sensitive indicator of key-elements in client and therapist interaction ... vocal measures are flexible enough to reflect moment-to-moment shifts in participation as well as more enduring styles.* (p. 74). Nevertheless, empirical research aimed at investigating the relationship between the psychotherapeutic process and speech prosody is still sparse (Horowitz, Sampson, Siegelman, Wolfson, & Weiss, 1975; Roessler & Lester, 1976; Rice, Koke, Greenberg, & Wagstaff, 1979; Wiseman & Rice, 1989; Levitt, 2001; Frankel, Levitt, Murray, Greenberg, & Angus, 2006; Campanelli, Ibern, Sarracino, Degni, & Mariani, 2007; Tomicic, 2011). This lack of empirical investigation is even more noticeable when specifically considering the ROS.

We were able to locate only a few studies (Spivack, 1996; Rocco, 2005, 2008; Rocco, Mariani, & Zanelli, 2013) that investigated the relationship between the therapeutic process and ROS. One of these studies (Spivack, 1996) investigated the degree of attunement in the ROS (assessed as words spoken in a 10-second interval) of the psychotherapeutic dyad in two long-term psychoanalytic treatments (good vs poor outcome). Attunement was analyzed using a time-series analysis to establish the degree to which the patient's ROS at one point in time influenced the therapist's ROS at a later point and vice-versa. The study results showed mutual influence only in the good outcome treatment, supporting claims that ROS attunement is correlated with clinically relevant therapeutic processes. A similar result was obtained by Rocco (2008), who investigated the paraverbal attunement between clients and therapists in terms of the ROS correlation in two brief psychodynamic psychotherapies (good vs poor outcome). A sample of three sessions was extracted from each treatment. The ROS was assessed in terms of syllables per sec-

ond. It was possible to observe a significant correlation between the client's and therapist's ROS in both treatments; however, the good outcome treatment presented significantly higher values. In another study, Rocco (2005) investigated the relationship between clients' ROS and their ability to link verbal and non-verbal representations in a single case of brief psychodynamic psychotherapy. The former was measured in terms of syllables per second, while the latter was assessed through the assessment of referential activity (RA), a measure of the extent to which subsymbolic information processing may be translated into symbolic information processing (Bucci, 1997). According to Bucci's multiple code theory (1997), subsymbolic information processing mainly regards bodily states, while symbolic information processing regards mental images and language. Results showed that ROS has a significant negative correlation with RA; thus, the more the connections activated by the client between the sub-symbolic and the symbolic system, the lower the speaking rate. Analogous results were obtained by Rocco et al. (2013).

The Therapeutic Cycles Model: emotional-cognitive regulation in the therapeutic process

The TCM (Mergenthaler, 1996, 2008, 2015) is a model of the therapeutic process that allows for its analysis in terms of the emotional-cognitive regulation in the dyad (client and/or therapist). Emotional-cognitive regulation is conceptualized as resulting from the interaction of *emotional* and *abstract-reflective* processes, representing two different modalities of information processing (Bucci, 1997), and characterizes the psychotherapeutic process of affective experiencing and reflection (see also Fonagy, Gergely, Jurist, & Target, 2005).

The TCM makes the assumption that these emotional and abstract-reflective processes (P_{ET} and P_{AB} , respectively) are represented at the level of discourse in the form of specific linguistic markers, namely, certain words are consid-

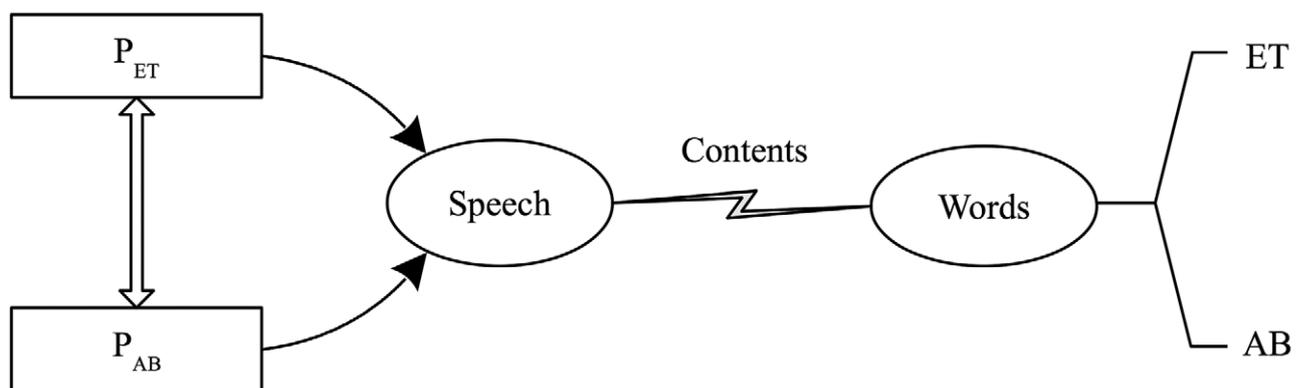


Figure 1. Model of speech production as a result of the interaction of emotional and abstract-reflective processes. P_{ET} , emotional processes; P_{AB} , abstract-reflective processes; ET, emotional words; AB, abstract words.

ered to mark the activation of an underlying psychological process (Figure 1). P_{ET} is operationalized as *emotional tone* (ET), which, in a therapeutic transcript, measures the frequency of words with an emotional connotation (Sandhöfer-Sixel, 1988), while P_{AB} is operationalized as *abstraction* (AB), which indicates the frequency of words considered to represent the verbal expression of abstract-reflective thought (Piaget, 1977; Schneider, 1983).

To measure the amount of ET and AB in the therapeutic dialogue, a computer-assisted deductive content analysis (Mayring, 2000) is carried out on the basis of one ET dictionary (containing emotional words) and one AB dictionary (containing abstract words). A computer program measures the frequency of ET and AB in units of analysis with a predefined size [usually word-blocks (WBs) for micro-analysis and/or entire sessions for macro-analysis] within the transcripts. These frequencies express the extent to which P_{ET} and P_{AB} are active during the therapeutic discourse within each unit of analysis. A higher activation of each of these two processes corresponds to a higher amount of ET and AB (and vice-versa). Figure 1 depicts the relationship between P_{ET} and P_{AB} active in the speaker during therapy and the speaker's emotional and abstract language.

The TCM further assumes that both P_{ET} and P_{AB} and, consequently, ET and AB words undergo a constant variation throughout the treatment. To assess this variation, the software standardizes the values of both ET and AB for each WB within a session transcript. On the basis of this transformation, the TCM defines four emotion abstraction patterns (EAPs), each representing one specific modality of emotional-cognitive regulation (Mergenthaler, 1996, 2008, 2015; Figure 2).

Relaxing (low emotion; low abstraction). Generic topics not characterized by emotional or conceptual contents (*i.e.*, *generic* emotional-cognitive regulation). In the client, it may correspond to introductory elements of the discourse and, in some cases, alexithymic narrative

modalities. In the therapist it may instead indicate moments of preliminary work, aimed primarily at interpersonal coordination.

Reflecting (low emotion; high abstraction). Discourse with high levels of abstraction (*i.e.*, *reflective* emotional-cognitive regulation). In the patient it may indicate the activation of reflective processes of reflection and effective coping, self-enhancement bias or intellectualizing defenses. In the therapist it may indicate interventions aimed at explaining inner states and/or the state of affairs.

Experiencing (high emotion; low abstraction). Verbalization of emotional content (*i.e.*, *experiential* emotional-cognitive regulation). It may signal the client's access to his/her emotional content, while in the therapist it may indicate self-disclosure and interventions aimed at facilitating the client's emotional awareness.

Connecting (high emotion; high abstraction). The speakers are able to verbalize emotional contents and are also able to reflect on them (*i.e.*, *integrative* emotional-cognitive regulation), reflecting key moments during the therapeutic *working through* or therapeutic work, for example, emotional insight, emotional-cognitive integration, and similar concepts. In the client it may correspond to moments of emotional insight, working through, and emotional-cognitive restructuring; in the therapist it may indicate the use of challenging (rather than supportive) interventions.

A final assumption of the TCM is that during clinically relevant moments in the course of therapy, the different EAPs follow a prototypical sequence over time: relaxing → experiencing → connecting → reflecting → relaxing (Figure 2). This temporal sequence is called the *therapeutic cycle*, which ideally represents a moment of heightened and productive clinical engagement (Mergenthaler, 1996, 2008).

The validity of the TCM with regard to the EAPs (and their organization in therapeutic cycles) has been supported by several studies, showing that the TCM allows for a linguistic-based assessment of emotional-cognitive

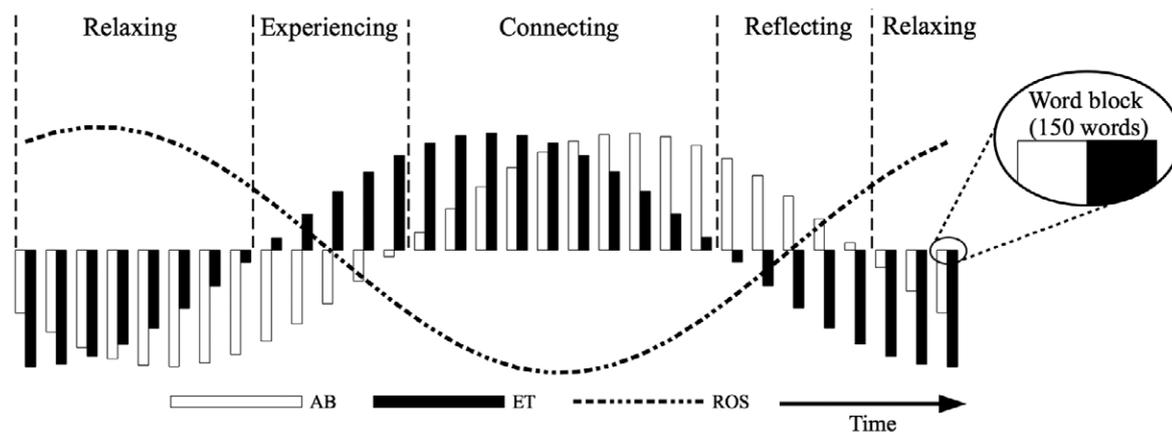


Figure 2. Relationship between the prototypical sequence of the different emotion-abstraction patterns and the rate of speech. AB, abstract words; ET, emotional words; ROS, rate of speech. In the projection, a word-block is presented.

regulation and, consequently, key moments of therapeutic work (e.g., Buchheim & Mergenthaler, 2000; Mergenthaler, 2008; Gelo & Mergenthaler, 2012).

Aims and rationale of the study

The present study aims to provide preliminary empirical evidence for the relationship between a client's emotional-cognitive regulation, as defined by the TCM, and his/her ROS within the therapeutic process.

Assumptions

First, we rely on a general assumption: the client's linguistic and paralinguistic behavior (measured by the TCM's variables and ROS, respectively) are both the product of his/her underlying P_{ET} and P_{AB} . Therefore, we expect each channel (linguistic or paralinguistic) to convey information on these underlying processes. Within the context of this general assumption, we further rely on a series of more specific assumptions as well as on empirical findings regarding: i) speech production and ROS; ii) emotional-cognitive regulation in the therapeutic process as described by the TCM, and, consequently; iii) the relationship between ROS and emotional-cognitive regulation.

Speech production and rate of speech

Speech production can be considered a complex information-processing procedure during which speakers transform intentions, thoughts, and feelings into fluently articulated speech (Levelt, 1993). The way information is processed along this path can be influenced by factors that are inherent to the environment and task as well as to the individual (e.g., attentional and memory resources).

Moreover, according to the *resource theory* (Kahneman, 1973; Wickens, 1980, 1992; Sweller, van Merriënboer, & Paas, 1998; Magill, 2001), the process of speech production is characterized by a limited amount of resources available to the individual in terms of attentional and memory capacities. Following this theory, the more complex or difficult the operations to be executed during the process of speech production, the greater the *cognitive load* and outlay of effort (Sweller et al., 1998).

Finally, paraverbal cues have been found to be sensitive indicators of a subject's cognitive load (Brenner, Roder, Hodel, Kienzie, Reed, & Liberman, 1994). More specifically, previous research has shown that higher levels of cognitive load result in a significantly lower ROS (Oviatt, Coulston, & Lunsford, 2004; Jameson, Kiefer, Muller, Großmann-Hutter, Wittig, & Rummer, 2010).

Emotional-cognitive regulation and the Therapeutic Cycles Model

First, the degree of activation of P_{ET} and P_{AB} is reflected, at a *content* level, in the frequency of ET and AB words (Figure 1). Second, P_{ET} and P_{AB} are two fundamentally independent yet interacting processes that have dif-

ferent combinations that result in four modalities of emotional-cognitive regulation (EAPs): *relaxing*, *reflecting*, *experiencing* and *connecting* (Figure 2). Each of these EAPs is characterized by different amounts of ET and AB (see the previous section about the TCM), which indicate the overall activation of P_{ET} and P_{AB} together. *Relaxing* reflects a low overall activation (ET and AB are both low), *reflecting* and *experiencing* represent a medium overall activation (one process is low and the other is high), and *connecting* reflects a high overall activation (ET and AB are both high). Third, according to the resource theory, higher overall activation of P_{ET} and P_{AB} causes a higher cognitive load for the speaker.

Relationship between rate of speech and emotional-cognitive regulation

The extent of the activation of a client's P_{ET} and P_{AB} is reflected in his/her speech not only at the *content* level but also at the *prosodic* level. In other words, the effect of the concurrent and interacting activation of these two processes can be traced to their epiphenomenal effect on the prosodic aspects of communication, namely the ROS. More specifically, in agreement with the resource theory, the degree of activation of a client's P_{ET} and P_{AB} is negatively correlated with his/her ROS.

Hypotheses

Based on the above assumptions, the following hypotheses can be formulated (see the Methods section for the definitions of measures).

First, the degrees of activation of P_{ET} and P_{AB} are proportional to the frequencies of ET and AB and are inversely proportional to ROS. Therefore, we expect a significant negative correlation between the sum of the ET and AB frequencies and the ROS (hypothesis 1).

Second, *relaxing* contains the least overall amount of ET and AB, *reflecting* and *experiencing* contain an average overall amount of ET and AB, and *connecting* contains the highest overall amount of ET and AB. Thus, we expect that ROS will be different in the following three distinct EAP conditions: *connecting*, *reflecting/experiencing* and *relaxing*. More specifically, we expect ROS will show the highest values in *relaxing*, medium values in *reflecting/experiencing* and lowest values in *connecting* (hypothesis 2; Figure 2).

Materials and Methods

Sample

The text analyzed is session 4 out of 28 sessions of an individual short-term psychodynamic psychotherapy treatment (Lewis, Dennerstein, & Gibbs, 2008) of a 36-year-old woman diagnosed with bulimia nervosa on Axis I and borderline personality disorder on Axis II [Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV-TR;

American Psychiatric Association, 2000]. The therapist, a male, was an experienced clinician with training in short-term psychodynamic psychotherapy. The sessions took place once a week in a *vis-à-vis* setting. The treatment was considered as having a good outcome according to the clinical judgment of the therapist; no measures regarding the outcome were available. The therapy was carried out in Italian. The client gave informed consent to the audio recording and transcriptions of the sessions.

This session was selected because the level of interruptions by the client while the therapist was speaking (or vice-versa) was lower than in other sessions. This greatly simplified the technical procedure of alignment (see following section on ROS). The session was tape-recorded and transcribed following the Italian adaptation (Mergenthaler, Freni, Giampieri, & Ferrari, 1998) of international psychotherapy transcription standards (Mergenthaler & Stinson, 1992).

Procedures and measures

The procedure described below was implemented to calculate the TCM and ROS values and to make the two datasets comparable.

Therapeutic Cycles Model measures: client's emotional and abstract language as well as his/her emotion-abstraction patterns

The whole session was analyzed to measure the following language measures: the client's ET and AB language as well as his/her EAPs. TCM linguistic measures are usually measured through the *cycle model* program (Mergenthaler, 1996) with reference to units of analysis: WBs of 150 words for the English, Italian and Spanish languages or 200 words for the German language (Mergenthaler, 1996, 2008). In this case, however, due to the specific requirements of this study, the TCM procedure was re-implemented in the *ad-hoc* program hABETROS (Tonti, 2005).

The two existing dictionaries of ET and AB for the Italian language have been adapted for hABETROS. The ET dictionary contained adjectives (*e.g.*, the Italian words for nice, wonderful, sad, bad, evil), verbs (*e.g.*, the Italian words for honor, abandon, love, hate) and adverbs (*e.g.*, the Italian words for alone, gladly); all are considered to express an emotional meaning according to the following dimensions: pleasure-displeasure, approval-disapproval, attachment-disattachment, and surprise (Sandhöfer-Sixel, 1988). The AB dictionary contained abstract nouns (*e.g.*, the Italian words for beauty, hate, plan, trial, advice), which are considered to be the verbal expression of abstract-reflective thought (Piaget, 1977; Schneider, 1983).

The textual analysis was performed by counting the number of ET and AB words in each block of 150 words [according to Mergenthaler (1985) the minimal text size for the English, Spanish, and Italian language is 150 words]; given that the last block was shorter than the 150

words required, it was added to the previous block (according to the standard TCM procedure). Then, the relative frequencies of ET and AB for each WB were transformed into *z*-scores using mean (*M*) and standard deviation (*SD*) from the whole session. Finally, these values were used to determine the EAPs by setting the following thresholds: *relaxing*: $z(ET) \leq 0$, $z(AB) \leq 0$; *reflecting*: $z(ET) \leq 0$, $z(AB) > 0$; *experiencing*: $z(ET) > 0$, $z(AB) \leq 0$; and *connecting*: $z(ET) > 0$, $z(AB) > 0$. As a result, it was possible to determine which EAP was displayed during every single WB comprising the session.

Rate of speech measures

The ROS was measured by using a lexical-based approach. Lexical-based approaches consist of counting the amount of a particular lexical unit (*e.g.*, characters, syllables, or words) per unit of time. In the present study, ROS was measured in words per second (WPS).

Alignment of the transcript

To measure the ROS, the textual information present in the transcript needs to be aligned with the information present in the audio recording; in other words, the text has to become a *timed* transcript. Ideally, every word's time boundaries should be identified, but this is sometimes very difficult to achieve. In our study, the client in the session analyzed presents with a rather high ROS (ranging from 174 to 234 words per minute on a WB basis, *i.e.*, about 3 to 4 WPS on average; *this is a personal evaluation of the authors, since no similar data on other clients could be located in the literature, however, 3 or 4 WPS seem actually to represent a rather high ROS*), which, together with the low quality recording, made it difficult to segment the session on a word level. To overcome this practical limit, we decided to proceed by segmenting the transcript into small text chunks (see below for details) and measuring the temporal boundaries of every chunk. We called these chunks *segments*. The result of this process of segmentation is described in Figure 3.

As we were interested in the client's speaking rate, we took into account just her speech, excluding the therapist's speech. Only the verbal speech (*i.e.*, the words) of the client has been considered, excluding silences and paraverbal cues, such as *m-hm* and *mmh?* which were already coded in the transcript (Figure 3) and therefore easily bypassed programmatically on the basis of a stop-list of *non-words*. Regarding silences, while it is true that in some cases they can be considered (fully or possibly partially) as part of the elaboration of the sentence being spoken, it is not possible to clearly categorize them on this basis; therefore, we decided to completely remove them from the computation. Finally, portions of the transcripts with overlapping client's and therapist's speech were excluded because the speech of the therapist (and the interruption itself) is assumed to influence the speaking rate of the

client if the two are talking at the same time. The program used for the segmentation of the text is Transcriber (Baras, Geoffrois, Wu, & Liberman, 2001), originally conceived to transcribe radio broadcasts. The program saved the data in a standard XML-formatted file, which was then imported into hABETROS (Tonti, 2005).

The criteria for the segmentation were: i) the segment had to be as short as possible, considering 0.5 seconds as an acceptable lower threshold; ii) no word could be truncated. This resulted in 1134 verbal segments containing 1 to 6 words ($M=3.34$; $SD=1.65$) and lasting from 0.5 to 2 seconds ($M=0.98$; $SD=0.31$). On the basis of this segmentation (Figure 3), by taking the segmentation made using Transcriber as input, hABETROS was able to identify the duration and textual content of each segment and then determine each segment's ROS. The ROS was computed by dividing the number of words contained in the segment

by the segment's duration. To compute the ROS over a longer period of time (e.g., the time required to speak the 150 words in a WB), the overall number of words within the period is divided by the actual time required to speak them, after filtering out the silent, overlapping, and therapist's segments.

Conjoined elaboration of Therapeutic Cycles Model measures and rate of speech

As final step in our procedure, we had to allow for a conjoined elaboration of TCM measures and ROS. Because ROS can be computed solely on the basis of spoken segments, the computation of values pertaining to TCM must be adapted to this data structure. To achieve this, the method of building the WBs must be modified. In the TCM the *building blocks* of WBs are single words, taken incrementally from the transcript until the minimal

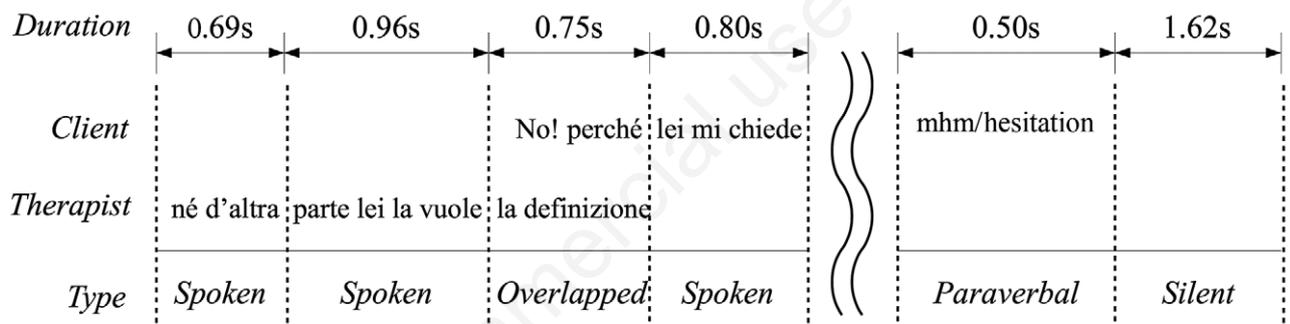


Figure 3. Part of a transcript divided into segments enriched with the timing information of each segment [duration expressed in seconds (s)], the client's and therapist's transcribed speech, and the type of segment.

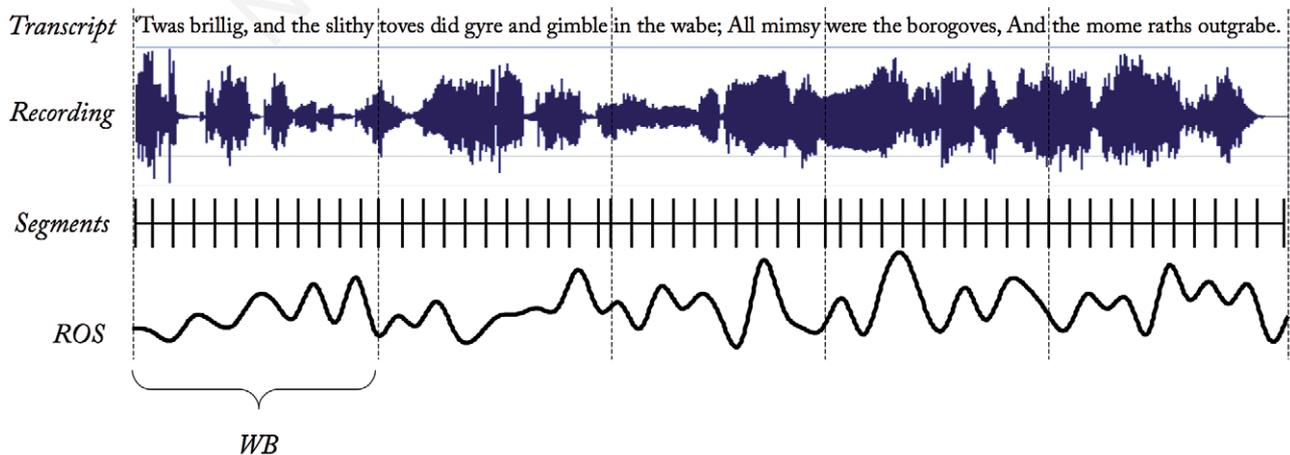


Figure 4. Exemplified representation of the relation between the text in the transcript (represented here with *Jabberwocky* by Lewis Carroll), the recording, the timed segments of text, the rate of speech (ROS) for each segment, and the word block (WB) from the therapeutic cycles model.

WB size of 150 words is reached (Figure 4). For the computation carried out in this research, the *building blocks* were the *segments* (see the subsection *Alignment of the transcript* above), which contain one or more words and at the same time convey the data about the timing and the type of vocal emission or silence. Variations in the WB sizes due to the fact that segments sometimes can contain more than one word are compensated statistically by computing z-scores on the relative frequencies (which take into account the size of the blocks) rather than directly analyzing the amounts of ET and AB words belonging to a WB.

This procedure allows for the extraction of data from the prosodic information and TCM compositionally in the same unit of analysis (the WB), making it possible to implement the desired statistical investigations.

Statistical analysis

Parametric tests were used because the data were normally distributed, as verified with a Shapiro-Wilk test value for the ROS [$W(26)=.976$, not significant (ns)], and for ET+AB [$W(26)=.974$, ns]. The first hypothesis of a negative proportionality between ET+AB and the ROS was assessed through a Pearson's correlation between ET+AB and the ROS.

The second hypothesis of the highest ROS values in *relaxing*, medium ROS values in *reflecting* and *experiencing*, and lowest ROS values in *connecting* was tested by comparing the means of the ROS across these three conditions with an ANOVA followed by a Dunnett *post-hoc* analysis.

Results

Hypothesis 1

The two data series used to test our first hypothesis were obtained based on the values from each WB (N=26). The first data series represents the overall activation defined as the sum of the ET and AB values (ET+AB) for each WB. The second data series is the client's ROS for each WB. It was possible to observe a highly significant negative correlation between the two data series [1-tailed Pearson's correlation: $r(26)=-.577$; $P<.001$], confirming our first hypothesis. Figure 5 displays the two data series being considered, where the ROS has been converted to z-scored values for comparison. It is possible to observe how the two data series negatively correlate with each other, with negative correlation peaks like, for example, in WBs 7 and 13. It is also possible to observe a few instances of positive correlation, such as for example in WBs 6 and 12. It is worth noting that lowest peaks in the ROS correspond to the highest peaks in the ET+AB and that those two WBs are of the *connecting* type (Table 1).

Hypothesis 2

The WBs of the analyzed session were grouped into the three EAP conditions reflecting the level of the expected ROS: high (in *relaxing*), medium (in *reflecting* and *experiencing*), and low (in *connecting*). Table 1 shows each WBs (N=26) EAP, expected level of ROS, and actual measured ROS.

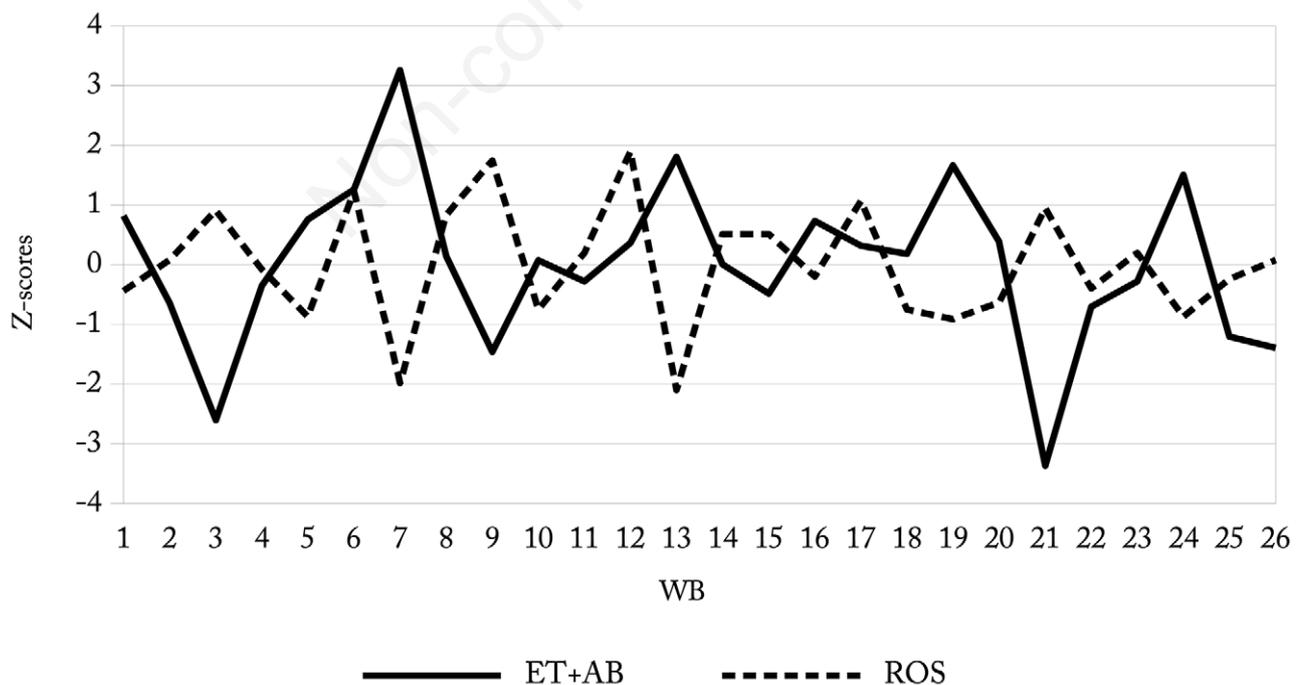


Figure 5. Plot of emotional and abstract (ET+AB) words and rate of speech (ROS) values for each word block (WB) composing the session. Both of the values are standardized for ease of comparison.

The descriptive statistics for the ROS in each of the three conditions are shown in Table 2. It was possible to observe a highly significant difference in the ROS among the three conditions [$F(2, 25)=7.339$, $P=.003$; Levene's test of homogeneity of variance=.917, ns]. The average ROS in the three conditions were then compared pairwise using *post-hoc* Dunnett tests (Table 2). The average ROS for the three conditions was in the order hypothesized: *relaxing* (high)>*reflecting/experiencing* (medium)>*connecting* (low).

Hypothesis 2 can be confirmed, although only partly. In fact, the ROS for the two EAP conditions – *relaxing* and *reflecting/experiencing* – differed significantly from the ROS for the *connecting* condition. Because the Dunnett test is directional, this confirms the fact that during the *connecting* condition words are effectively spoken with a lower ROS compared to the other EAP conditions. However, the hypothesis of *reflecting/experiencing* having a lower ROS than *relaxing* cannot be fully confirmed.

Discussion

The goal of this study was to preliminarily investigate the relationship between a client's emotional-cognitive regulation, as defined by the TCM, and ROS within the therapeutic process. Our general hypothesis was that there would be a relationship between the client's linguistic behavior (in terms of ET, AB, and their different combinations in EAPs) and prosodic behavior (in terms of ROS), mediated by the degree of activation of P_{ET} and P_{AB} . More specifically, we expected that the client's ROS would be inversely proportional to the sum of ET and AB (hypothesis 1). Moreover, we expected that different EAPs would be characterized by a different overall level of activation, consequently showing a different ROS, with the highest values in *relaxing*, medium values in *reflecting/experiencing*, and lowest values in *connecting* (hypothesis 2).

From our study data, it was possible to support our first hypothesis of a negative correlation between the sum of ET and AB and the ROS. In other words, a higher overall amount of ET and AB spoken by the client correlates with a lower ROS (and vice-versa). This result is consistent with our assumption based on the resource theory (Kahneman, 1973; Wickens, 1980, 1992; Magill, 2001), suggesting that a higher (or lower) degree of activation of P_{ET} and P_{AB} , resulting in a higher (or lower) amount of ET and AB, effec-

tively represents a higher (or lower) cognitive load for the client's information-processing system, consequently decreasing (or increasing) the ROS proportionally.

Our second hypothesis could only be partially con-

Table 1. Emotion-abstraction patterns, and expected and measured rate of speech level for each analyzed session's word block.

| WB | EAP | Expected ROS | ROS (WPS) |
|----|---------------------|--------------|-----------|
| 1 | <i>Connecting</i> | Low | 3.32 |
| 2 | <i>Reflecting</i> | Medium | 3.45 |
| 3 | <i>Relaxing</i> | High | 3.66 |
| 4 | <i>Reflecting</i> | Medium | 3.41 |
| 5 | <i>Reflecting</i> | Medium | 3.21 |
| 6 | <i>Reflecting</i> | Medium | 3.74 |
| 7 | <i>Connecting</i> | Low | 2.93 |
| 8 | <i>Reflecting</i> | Medium | 3.64 |
| 9 | <i>Relaxing</i> | High | 3.87 |
| 10 | <i>Reflecting</i> | Medium | 3.24 |
| 11 | <i>Reflecting</i> | Medium | 3.48 |
| 12 | <i>Experiencing</i> | Medium | 3.91 |
| 13 | <i>Connecting</i> | Low | 2.90 |
| 14 | <i>Experiencing</i> | Medium | 3.56 |
| 15 | <i>Reflecting</i> | Medium | 3.56 |
| 16 | <i>Connecting</i> | Low | 3.38 |
| 17 | <i>Experiencing</i> | Medium | 3.70 |
| 18 | <i>Experiencing</i> | Medium | 3.24 |
| 19 | <i>Experiencing</i> | Medium | 3.20 |
| 20 | <i>Connecting</i> | Low | 3.27 |
| 21 | <i>Relaxing</i> | High | 3.67 |
| 22 | <i>Reflecting</i> | Medium | 3.33 |
| 23 | <i>Reflecting</i> | Medium | 3.48 |
| 24 | <i>Connecting</i> | Low | 3.21 |
| 25 | <i>Relaxing</i> | High | 3.37 |
| 26 | <i>Experiencing</i> | Medium | 3.45 |

WB, word block; EAP, emotion-abstraction pattern; ROS, rate of speech; WPS, words per second.

Table 2. Average rate of speech in the three emotion-abstraction pattern conditions and *post-hoc* comparisons.

| EAP condition (expected ROS) ^o | | | Dunnett test (P) | | |
|---|--------------------------------|------------------|----------------------------------|---------------------|------------------------------------|
| Relaxing (n=4) | Reflecting/experiencing (n=16) | Connecting (n=6) | Relaxing>reflecting/experiencing | Relaxing>connecting | Reflecting/experiencing>connecting |
| 3.64 (0.21) | 3.47 (0.21) | 3.17 (0.2) | .128 | .002 | .004 |

EAP, emotion-abstraction pattern; ROS, rate of speech. ^oValues are expressed as mean (standard deviation).

firmed. As expected, the *connecting* EAP showed a significantly lower ROS compared to both of the other conditions (*relaxing* and *reflecting/experiencing*). This result suggests that when both P_{ET} and P_{AB} are active above average levels (*i.e.*, high overall activation, as in *connecting*), the client's cognitive load is significantly higher than when one of these processes is active above average levels and the other is active below average levels (*i.e.*, medium overall activation, as in *reflecting* or *experiencing*) or when both processes are active below average levels (*i.e.*, low overall activation, as in *relaxing*). However, the hypothesis of *reflecting/experiencing* showing significantly lower ROS values than *relaxing* could not be confirmed; the client's ROS significantly decreased only when shifting to a *connecting* state while, even if the ROS decreased in the expected direction (from 3.64 to 3.47), this change is not significant for the other conditions. This result suggests that when either P_{ET} or P_{AB} is active above average levels (*i.e.*, medium overall activation, as in *reflecting* or *experiencing*), the cognitive load for the client is not significantly higher than when both of these processes are active below average levels (*i.e.*, low overall activation, as in *relaxing*). Thus, it would seem that the client's cognitive load significantly increases *only* when shifting to a *connecting* state and remains the same in all other cases. One possible reason for this may be that *connecting*, characterized by integrative processes, requires additional cognitive load to be performed compared to *relaxing* and *reflecting/experiencing*. However, this result might be due to the small size of the sample and further investigations could shed light on the reasons of this partial result.

In sum, these findings are coherent with the empirical literature showing that *connecting*, compared to the other EAPs, plays a *key role* in the client's change process (Mergenthaler, 2008). More specifically, these results suggest that a significant decrease in the client's ROS may be a marker for *integrative* emotional-cognitive regulation, which has been shown to be an essential constituent of meaningful in-session clinical work and to be predictive of a good therapeutic outcome.

Limitations and suggestions for future studies

These results, while promising, must be carefully considered because of several limitations. First, because of its explorative nature, this study was based on a single session of psychotherapy, reducing the external validity and generalizability of the results (see Lepper & Mergenthaler, 2005 for a similar approach making use of a single session). Future studies should be conducted with a larger sample of sessions and treatments. It would be particularly interesting to collect data on both good and poor outcome therapies to explore if the supposed relationship between ROS and linguistic behavior is different with regard to the clinical outcome of treatments.

Second, because of the explorative nature of this

study, the analyses were based exclusively on the client's speech; this decision was taken to simplify the measurement of the ROS. Future studies should focus on both the client's and therapist's speech. This would allow for the investigation of possible relational attunements that client and therapist might develop over different phases of a treatment, and whether this attunement is related to the clinical outcome of the treatment. Previous research (Spivack, 1996; Rocco, 2005, 2008) has shown that clients and therapists tend to develop an interpersonal attunement in their ROS and that this attunement is higher in good-outcome therapies. Moreover, it would be interesting to investigate the time-lagged relationship between the therapist's and client's ROS; this would allow researchers to identify the extent to which the intentional modulation of the therapist's ROS might influence the ROS of the client, consequently affecting his/her emotional-cognitive regulation.

Third, the parts of the transcripts where the client's and therapist's speech overlapped were excluded not only from the measurement of the ROS (a limitation that was technically unavoidable) but also from the computation of the TCM values. Thus, some of the client's linguistic behavior has not been assessed in this study. However, because the amount of these words was extremely low, they can be considered as not relevant. Future studies should include the words spoken in overlapping segments in the formation of the TCM WBs. Still, the ROS of the WB should be measured on the basis of the plainly spoken segments only.

Fourth, the way that we calculated the different patterns of emotional-cognitive regulation (see Materials and Methods section) allowed us to assess if P_{ET} and P_{AB} were both active below average levels (*relaxing*), one was above and the other below average levels (*reflecting* and *experiencing*), or both were above average levels (*connecting*). However, we did not assess *how much* each of the two processes was active below or above mean levels. Future studies should address this issue, for example, by setting a computation threshold in terms of standard deviation from the mean.

Fifth, we did not assess the extent to which the length of different words might affect the ROS of the client. Since it might be supposed that the different EAPs are characterized by words with a different average length (*e.g.*, *reflecting* contains abstract words, which may tend to be longer, while *relaxing* may contain function words, like articles, pronouns, conjunctions, which tend to be shorter), future studies should control for this.

Finally, following the TCM standard for the Italian, English, and Spanish languages, the analyses were performed in WBs of 150 words. Because the ROS may vary greatly from second to second, it may be hypothesized that assessing both the TCM linguistic behavior and ROS within WBs of different sizes may influence the relationship between the two. Future studies should try to replicate the observed results in WBs of different sizes.

Conclusions

One of the essential features of psychotherapy is that the therapist must be able to make meaningful inferences about the inner processes of the client by what he/she communicates during the sessions. This ability is essential for the therapist to monitor and regulate the therapeutic process. Research with the TCM has shown that specific patterns of a client's *linguistic* communication (in terms of the frequency of uttered ET and AB words) represent sensitive and reliable indicators of different modalities of emotional-cognitive regulation (*relaxing, reflecting, experiencing* and *connecting*). The results presented in this paper provide preliminary empirical support to the idea that the features of client's *paralinguistic* communication (namely, his/her ROS) may allow the therapist to make meaningful inferences concerning the different modalities of emotional-cognitive regulation that the client is engaged with during the course of psychotherapy. More specifically, these results suggest that a significant reduction in the client's ROS is associated with integrative strategies of emotional-cognitive regulation, which previous empirical research has shown to be a necessary component of the client's change process.

These results, if further confirmed, may have relevant implications for clinical practice. Clinicians should learn to detect significant reductions in the client's ROS, which would then provide them with precious moment-to-moment information on when the client enters key moments of heightened therapeutic work, characterized by his/her ability to reflect on emotional contents. This piece of information, in turn, may be used by the clinician to monitor the state and progress of in-session work with regard to the client's emotional-cognitive regulation, consequently modulating the type and timing of the interventions performed on these regulative processes, and leading to feedbacks on the effects of these interventions. Clinical supervision represents an elective context where all these issues might be shared and discussed.

References

- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders (4th ed., Text Revision)*. Washington, DC: American Psychiatric Association.
- Bady, S.L. (1985). The voice as a curative factor in psychotherapy. *Psychoanalytic Review*, 72(3), 479-490.
- Bänninger-Huber, E., & Widmer, C. (1999). Affective relationship patterns and psychotherapeutic change. *Psychotherapy Research*, 9(1), 74-87. doi: org/10.1080/10503309912331332601
- Barras, C., Geoffrois, E., Wu, Z., & Liberman, M. (2001). Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication*, 33(1-2), 5-22. doi: org/10.1016/S0167-6393(00)00067-4
- Bemieri, F.J. (1989). Coordinated movement in human interaction: synchrony, posture similarity, and rapport. *Dissertation Abstracts International*, 49, 4601.
- Brenner, H., Roder, V., Hodel, B., Kienzie, N., Reed, D., & Liberman, R. (1994). *Integrated psychological therapy for schizophrenic patients*. Seattle, WA: Hogrefe & Huber.
- Bucci, W. (1997). *Psychoanalysis and cognitive science: a multiple code theory*. New York, NY: Guilford Press.
- Buchheim, A., & Mergenthaler, E. (2000). The relationship among attachment representation, emotion-abstraction patterns, and narrative style: a computer-based text analysis of the adult attachment interview. *Psychotherapy Research*, 10(4), 390-407. doi: org/10.1093/ptr/10.4.390
- Campanelli, L., Ibern, E., Sarracino, D., Degni, S., & Mariani, R. (2007). Semiotics of the non verbal vocal expression of emotions and research into the psychotherapy process: a pilot study. *Rivista di Psicologia Clinica*, 1, 102-115.
- Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2005). *Affect regulation, mentalization, and the development of self*. New York, NY: Other Press.
- Frankel, Z., Levitt, H.M., Murray, D.M., Greenberg, L.S., & Angus, L. (2006). Assessing silent processes in psychotherapy: an empirically derived categorization system and sampling strategy. *Psychotherapy Research*, 16(5), 627-638. doi: org/10.1080/10503300600591635
- Gelo, O.C.G., & Mergenthaler, E. (2012). Unconventional metaphors and emotional-cognitive regulation in a metacognitive interpersonal therapy. *Psychotherapy Research*, 22(2), 159-175. doi: org/10.1080/10503307.2011.629636
- Gelo, O.C.G., Salcuni, S., & Colli, A. (2012). Text analysis within quantitative and qualitative psychotherapy process research: introduction to special issue. *Research in Psychotherapy: Psychopathology, Process and Outcome*, 15(2), 45-53. doi: org/10.7411/RP.2012.005
- Horowitz, L.M., Sampson, H., Siegelman, E.Y., Wolfson, A., & Weiss, J. (1975). On the identification of warded-off mental contents: an empirical and methodological contribution. *Journal of Abnormal Psychology*, 84(5), 545-558. doi: org/10.1037/h0077139
- Jameson, A., Kiefer, J., Müller, C., Großmann-Hutter, B., Wittig, F., & Rummer, R. (2010). Assessment of a user's time pressure and cognitive load on the basis of features of speech. In: M. W. Crocker, & J. Siekmann (Eds.), *Resource-adaptive cognitive processes* (pp. 171-204). Berlin, Heidelberg: Springer. doi: org/10.1007/978-3-540-89408-7_9
- Jarmołowicz, E., Karpinski, M., Malisz, Z., & Szczyszek, M. (2007). Gesture, prosody and lexicon in task-oriented dialogues: multimedia corpus recording and labelling. In: A. Esposito, M. Faundez-Zanuy, E. Keller, & M. Marinaro (Eds.), *Verbal and nonverbal communication behaviours* (pp. 99-110). Berlin/Heidelberg: Springer. doi: org/10.1007/978-3-540-76442-7_9
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Keller, E., & Tschacher, W. (2007). Prosodic and gestural expression of interactional agreement. In: A. Esposito, M. Faundez-Zanuy, E. Keller, M. Marinaro (Eds.) *Verbal and nonverbal communication behaviours* (pp. 85-98). Berlin: Springer-Verlag.
- Lepper, G., & Mergenthaler, E. (2005). Exploring group process. *Psychotherapy Research*, 15(4), 433-444. doi: org/10.1080/10503300500091587
- Levelt, W.J.M. (1993). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

- Levitt, H.M. (2001). Sounds of silence in psychotherapy: the categorization of clients' pauses. *Psychotherapy Research, 11*(3), 295-309. doi: org/10.1080/713663985
- Lewis, A.J., Dennerstein, M., & Gibbs, P.M. (2008). Short-term psychodynamic psychotherapy: review of recent process and outcome studies. *The Australian and New Zealand Journal of Psychiatry, 42*(6), 445-455. doi: org/10.1080/00048670802050520
- Magill, R.A. (2001). *Motor learning: concepts and applications*. Boston, MA: McGraw-Hill.
- Mayring, P. (2000). Qualitative content analysis. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research, 1*(2), 20.
- Mergenthaler, E. (1985). *Textbank systems: computer science applied in the field of psychoanalysis*. Heidelberg: Springer-Verlag.
- Mergenthaler, E. (1996). Emotion-abstraction patterns in verbatim protocols: a new way of describing psychotherapeutic processes. *Journal of Consulting and Clinical Psychology, 64*(6), 1306-1315. doi: org/10.1037//0022-006X.64.6.1306
- Mergenthaler, E. (2008). Resonating minds: a school-independent theoretical conception and its empirical application to psychotherapeutic processes. *Psychotherapy Research, 18*(2), 109-126. doi: org/10.1080/10503300701883741
- Mergenthaler, E. (2015). Resonating minds: a school-independent theoretical conception and its empirical application to psychotherapeutic processes. In: B.M. Strauss, J.P. Barber, & L.G. Castonguay (Eds.), *Visions in psychotherapy research and practice. Reflections from presidents of the society for psychotherapy research* (pp. 292-314). New York, NY: Routledge.
- Mergenthaler, E., Freni, S., Giampieri, E., & Ferrari, R. (1998). Regole standardizzate di trascrizione delle sedute di psicoterapia [Standardized transcription rules of psychotherapy sessions]. *Ricerca in Psicoterapia, 1*, 29-46.
- Mergenthaler, E., & Stinson, C.H. (1992). Psychotherapy transcription standards. *Psychotherapy Research, 2*(2), 125-142. doi: org/10.1080/10503309212331332904
- Neumann, R., & Strack, F. (2000). "Mood contagion": the automatic transfer of mood between persons. *Journal of Personality and Social Psychology, 79*(2), 211-223. doi: org/10.1037//0022-3514.79.2.211
- Núñez, R.E., & Sweetser, E. (2006). With the future behind them: convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science, 30*(3), 401-450. doi: org/10.1207/s15516709cog0000_62
- Oviatt, S., Coulston, R., & Lunsford, R. (2004). When do we interact multimodally? In: *Proceedings of the 6th international conference on Multimodal interfaces - ICMI '04* (p. 129). New York, NY: ACM Press. doi: org/10.1145/1027933.1027957
- Piaget, J. (1977). *Recherches sur l'abstraction réfléchissante [Research on reflective abstraction]*. Paris: Presses Universitaires de France.
- Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology, 79*(3), 284-295. doi: org/10.1037/a0023419
- Rice, L., & Kerr, G. (1986). Measures of client and therapist vocal quality. In: L. Greenberg & W. Pinsof (Eds.), *The psychotherapeutic process. A research handbook* (pp. 73-105). New York, NY: Guilford Press.
- Rice, K., Koke, C., Greenberg, L., & Wagstaff, A. (1979). *Manual for client voice quality*. Toronto: York University Counselling and Development Centre.
- Rocco, D. (2005). Analisi degli aspetti paraverbali in una psicoterapia dinamica breve: il metodo dell'Attività Referenziale e l'analisi della Speech Rate [Analysis of the paraverbal aspects in a brief dynamic psychotherapy: the referential activity and the speech rate]. *Ricerca in Psicoterapia, 8*(1), 127-147.
- Rocco, D. (2008). Analisi empirica dell'influenzamento reciproco tra paziente e terapeuta. Il linguaggio dei contenuti ed il linguaggio delle emozioni [Empirical analysis of the reciprocal influence between patient and therapist.]. *Ricerca in Psicoterapia, 11*(1), 47-74.
- Rocco, D., Mariani, R., & Zanelli, D. (2013). The role of non-verbal interaction in a short-term psychotherapy: preliminary analysis and assessment of paralinguistic aspects. *Research in Psychotherapy: Psychopathology, Process and Outcome, 16*(1), 54-64. doi: 10.7411/RP.2013.007
- Roessler, R., & Lester, J.W. (1976). Voice predicts affect during psychotherapy. *The Journal of Nervous and Mental Disease, 163*(3), 166-176. doi: org/10.1097/00005053-197609000-00004
- Sandhöfer-Sixel, J. (1988). *Modalität und gesprochene Sprache: Ausdrucksformen subjektiver Bewertung in einem lokalen Substandard des Westmitteldeutschen [Modality and language: Expression modes of subjective attribution in a local sub-standard of west-middle German]*. Wiesbaden: Steiner Verlag.
- Schneider, H. (1983). *Auf dem Weg zu einem neuen Verständnis des Psychotherapeutischen Prozesses [On the way to a new understanding of the psychotherapeutic process]*. Bern: Huber.
- Spivack, N. (1996). *Measuring mutual influence in the analytic discourse*. Garden City, NY: Adelphi University.
- Sweller, J., van Merriënboer, J.J.G., & Paas, F.G.W.C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296. doi: org/10.1023/A:1022193728205
- Tomicic, A. (2011). *Vocal coordination: a window for the study of mutual regulation in psychotherapeutic dialog*. Santiago de Chile: Pontificia Universidad Católica de Chile.
- Tonti, M. (2005). hABETROS (ad-hoc computer program). A computational approach to the psychotherapeutic process. Available from: tonti.web.cs.unibo.it/etc/thesis.pdf
- Wennerstrom, A. (2001). *The music of everyday speech: prosody and discourse analysis*. Oxford: Oxford University Press.
- Wickens, C.D. (1980). The structure of processing resources. In: R. Nickerson (Ed.), *Attention and performance VII* (pp. 239-257). Hillsdale, NJ: Erlbaum.
- Wickens, C.D. (1992). *Engineering psychology and human performance*. New York, NY: HarperCollins.
- Wiseman, H., & Rice, L.N. (1989). Sequential analyses of therapist-client interaction during change events: a task-focused approach. *Journal of Consulting and Clinical Psychology, 57*(2), 281-286. doi: org/10.1037/0022-006X.57.2.281