Prosodic iconicity and experiential blending*

Antoine Auchlin
Geneva University

In order to account for prosodic iconicity in speech in a very general way
we propose looking at the phenomenon from an experiential and embodied
perspective (Núñez 1999; Violi 2003; Rohrer 2007, i.a.), defining communication
as a “co-experienciation” process. Using different paths, prosodic dimensions’
variations impose direct, non-mediated shaping of shared experience. Prosodic
iconic formations take place in that space of shared experience. The way it mixes
with meaning may be schematized using Fauconnier and Turner's Conceptual
Blending Theory. We suggest (following Hutchins 2005; Bache 2005) some
accommodation of the schema in order to take into account the perceptual
dimension of part of the blending input, as well as the experiential dimension
of blending output.

1. Introduction

The general concern of this proposal is to reconsider prosodic contribution
to the elaboration of speech events, and to communication. To put it meta-
phorically, following the “Jackendoff glasses” metaphor (Rohrer 2007): one can
take the glasses with the hand and hold them in front of oneself, examine and
describe them; when doing so, they are “objectified”, taken as an object. Another
use of glasses – the usual one – is to wear them and look through them. Then,
they are no longer objectified: they act as part of the perception process, and
are, in that way, incorporated – or “subjectivised”, no longer perceived as objects
of perception. Such is, in many respects, the phenomenal status of prosody, as
we scholars look at it: what we’re looking at is seen through. Consequently, the
interest here is to describe how prosody and prosodic iconic formations may
determine experience through perception.

* Thanks to Patrick Morency for his help with the English version; to Anne Catherine Simon
for valuable comments.
Conversely, to use Plato’s metaphor, looking for prosodic contribution internal to meaning formation only, in order to understand prosody, is like looking at shadows deep in the cave to understand the outer world. Prosody’s contribution to meaning is not exclusively within formed meaning. It is, or at least a large part of it is, outside of meaning, in the experience of meaning and its temporal emergence and unfolding as speech progresses.

Prosody and Iconicity meeting was intended to “(…) shed a new light on the interrelation between prosody and iconicity (…)”. The present paper aims at doing so through a few related assumptions.

Prosodic iconic formation involves motor icons and corporeal schemas, through direct experience duplication, as well as through (pre)motor imitation (developed below § 2.4).

Secondly, the import of prosodic iconic formation is in modeling speech experience itself, not only in the elaboration of an abstract conceptual interpretation (though not excluded): its use and function is to instantiate substantial experience of verbal content, hence creating some kind of stereophonic effect in meaning experience. The study of meaning as experience is relevant in an “embodied cognition” and Cognitive Linguistics general framework (Lakoff & Johnson 1985, 1999; Gibbs 2003; Rohrer 2007; Violi 2003, etc.).

We will consider the mutual contribution of prosody and verbal content within Fauconnier and Turner’s (2002) “Conceptual Blending” framework. Following (Hutchins 2005) we assume some accommodation of the framework, in order to take into account the specific material and perceptive dimension of part of the input, as well as the experiential dimension of blending output.

Indeed, not all prosodic manifestations are equally concerned by this proposal: there is a gradient between properly “grammaticalized” or conventionalized prosodic forms, intonational morphemes (Mertens 1987), on one hand, and raw phono-prosodic manifestations, like speech rate, or non-canonic formant position, timbre, on the other. This variety is considered a tripartition between natural signs, natural signals, and non-natural signals by Wilson and Wharton (2006; I comment on this further below). Non-natural, grammaticalized, forms may be thought to combine with linguistic forms in a compositional way. A high

1. “The purpose of the conference is to shed new light on the interrelation between prosody and iconicity by enlarging the number of parameters traditionally considered and by confronting various theoretical backgrounds.”

2. “The corporeal schema is not only the general kinaesthetic experience we have of our body, it is also the spatial dimension that is occupied by the body.” Violi (2003:215).

rise at the final edge of a proposition seemingly combines with propositional content in quite the same way as an interrogative form does (see French est-ce que), that is compositionally. Yet whatever prosodic morpheme is used, it goes through “phonetic implementation” (Gussenhoven 2002). Part of the information coded by the three hypothesized “biological codes” may have some importance (may contribute) to verbal content interpretation, but this contribution is not compositional; and whatever is the importance of this contribution, it does not exhaust phonetic implementation information: timbre, rhythm, speech rate are still necessary to experiencing speech (discussion of Wilson & Wharton 2006 proposal’s below, § 2).

Suppose I hit you: is my blow “coded”? Does it make any sense to think of it as something coded? Surely not: my punch directly determines your experience, in a direct way, not symbolically mediated or conceptualized. Such is the case of tempo, speech rate, and rhythm: it is hard to escape them. Whether they are further developed in interpretation or conceptual elaboration is another question.

Iconicity is a problematic matter only if it is considered from a (post-) saussurian point of view on language and its arbitrariness. But, as Meunier puts it:

> There are (then) good reasons, both theoretical an intuitive ones, to take seriously the Iconicity of Thought Hypothesis. If linguistic expression aims at building and communicating iconic mental models, there is no reason to subordinate it to a hypothetical propositional language of thought. (Meunier 2003: 111)

---

4. It is, or can be seen as, inferential, instead of compositional. Relevance Theory’s answer, schematically, is that phonetic details are (mutually) manifest to hearer and speaker, as “cognitive environments”; whatever is manifest may become relevant to interpretation, provided it is used as a premise in deriving implicatures from what is said (and other premises) (according Wilson & Wharton 2006). Hirschberg’s (2002) proposal is parallel to this; it uses a Gricean framework in order to describe how f0-coded information may get involved in utterances interpretation process. The Gricean framework however could be simplified following RT general lines on Gricean perspective, replacing, among other modifications, Grice’s four Conversation Maxims by a single general Relevance Principle.

5. As for generativists’ resistance against considering iconic processes in grammar, see Van Langendonck (2007: 396).

6. “On a (donc) de bonnes raisons, théoriques autant qu’intuitives, de prendre au sérieux l’hypothèse de l’iconicité de la pensée. (…) Si l’expression linguistique a pour but la construction et la communication de modèles mentaux de nature iconique, il n’y a aucune raison de la subordonner à un hypothétique langage de la pensée de nature strictement propositionnelle.” (Meunier 2003: 111).
2. The semiotic scene: Overt and global communication models

Wilson and Wharton (2006) propose a very elegant view of prosodic contributions to communication. Though it does not claim to be exhaustive, this account seems to include all possible prosodic facts and communicative importance, using two sets of distinctions. The distinction “between natural signs, natural signals, and non-natural signals (whether linguistic or cultural), Figure 1; and the distinction between accidental, covert, and overt information transmission, Figure 2 below (ibid: 1576).

![Diagram of prosodic inputs]

**Figure 1.** Varieties of prosodic inputs; Wilson and Wharton (2006: 1563)

![Diagram of prosodic information transmission]

**Figure 2.** Varieties of prosodic information transmission (Wilson & Wharton 2006: 1564)

However inclusive this account is, it restrains the scope of the facts to be considered. As a whole, the taxonomy does not consider the fact that all categories of prosodic signs and signals work together in communication, and that overt communication hardly goes without any accidental communication. Analytically, it does not account for raw acoustic determination of prosody, the facts, beyond “natural signals”, before they enter inference processes.

This is due to two distinct important hypotheses of Relevance Theory’s pragmatic framework: that communication may be studied from a hearer-only perspective; and that communication is abstract information (conceptual) processing.
2.1 The ‘hearer-only perspective’

Wilson & Wharton assume a ‘hearer-only perspective’;\(^7\) this perspective is distinct from the ‘speaker-perspective’, as in Gussenhoven (2002) – see discussion (\textit{ibid.}: 1569) about “effort”, as pertaining to speakers (Gussenhoven’s ‘effort code’) or hearers (Wilson & Wharton “cognitive processing effort”). More importantly, the “hearer-only” perspective is distinct as well from the ‘dance perspective’, including speaker and hearer, and their mutual adjustment. But interaction studies (Auer et al. 1999; Couper-Kühlen & Selting 1996; \textit{i.a.}) clearly show that mutual synchronization of action in interaction is fine-grained, monitored, and controlled prosodically – be it in an automated and transparent way, as in “normal” circumstances:

“The dance of nonverbal communication between two individuals often goes unnoticed by either participant (…). It is, however, noticeable when it is absent or out of sync.” (Lieberman 2000: 123; quote in Wilson & Wharton \textit{ibid.}, 1575)

Such transparency is the \textit{glasses metaphor} problem. It is correlated to the “abstract information processing” problem, below.

2.2 “Abstract information processing”

Wharton (2003, quoted in Wilson & Wharton \textit{ibid.}: 1562), illustrates the distinction between natural signs and natural signals by comparing shivering – a natural sign-, to smiling – a natural signal. Natural signs may provide evidence for something but are neither designed to, nor performed for, that use; natural signs are processed (by hearers) through inference, not decoding.

However what are called “natural signs” refer primarily to \textit{facts}, whose basic existence is not that of \textit{signs}. Facts become signs only by entering an inferential process. But what happens to those facts that do not pass through the “sign threshold”? They are \textit{not} considered as possible prosodic inputs by Wilson & Wharton (Figure 1). Yet, as we will illustrate, \textit{some} facts do not need to be a sign of anything to operate: temporal manifestations (speech rate, syllable duration, rhythm, etc.) operate in a \textit{direct} way, neither mediated by coding, nor by inferring: they operate as the temporal mutual tuning of speakers’ and hearers’ experiences.

Consequently, prosody should not exclusively be studied in a narrow framework centered on ‘overt communication’ and speaker meaning’s interpretation, as defined in the Gricean tradition. According to Wilson and Wharton (2006: 1564),

\footnote{Recovering the speaker’s informative and communicative intention is seen from the hearer’s point of view.}
“Relevance Theorists (…) have consistently argued that there is a continuum of cases between showing and meaning\textsubscript{NN}, all of which may fall within the domain of pragmatics and contribute to a speaker’s meaning” (Sperber & Wilson 1986/1995, Chapter 1, Section 10). (see Figure 3 below). But the continuum goes beyond showing to presenting, and the domain of pragmatics should not be restricted to focus only on speaker’s meaning and its recovery by hearers (readers).

![Overt Communication Continuum Diagram](image)

**Figure 3.** Varieties of overt communication and the showing-meaning\textsubscript{NN} continuum (Wilson & Wharton 2006: 1564)

A written example will allow discussion of this point:\textsuperscript{9}

(1) **Ensembles nous irons plus loin** (*togethers we’ll go further*)

Example (1) is the credo for an NGO trying to recruit new members. The point is in the plural mark added to invariable adverb (underlined). On one hand, it clearly indicates the writer’s desire for numerous people to join in; there is some quantity blending operation here. But on the other hand, the extra “s” acts as a symptom of a writer with incorrect spelling. Indeed, if the writer may assume the quantity blend, he surely does not assume the faulty form, which is accidentally or un-intentionally, made manifest.\textsuperscript{10} The faulty form is not part of overt communication. For the reader (some readers), it indeed is part of the speech event, and probably an important one: it can, among other things, shed some discredit on the exposed content, exhortation to join in, and so on. It’s then a case of a bad blend, communicative un-happiness. It may also undergo an integration process in the reader’s experience, through a complex elaboration that takes into account these different aspects. The extra “s” may be detected as faulty, identified as the writer’s enthusiasm at having many new people join, which overcomes his vigilance, a trait that typifies enthusiastic young people. This characterization of a text’s author may blend in turn as an image of the whole enterprise.

\textsuperscript{8} Meaning\textsubscript{NN}, or Non Natural Meaning, in the Gricean and “post-Gricean” framework, is speaker meaning conveyed by non natural signals in ostensive communication.

\textsuperscript{9} See Stroumza and Auchlin (1997); Auchlin and Burger (2008) for case-analyses on similar topics.

\textsuperscript{10} It is rendered manifest for some readers – but not all. Lots of people in the editing process could have seen it – but did not.
This is why the study of communication should not be restricted to overt communication. In order for overt communication to take place, in order for a piece of linguistic (or other) stimulus for communication to be shown, there must be some presenting of the stimulus. Equating what is presented to what is shown, or ignoring the crucial difference between them, is an idealization, an over-simplification that leads to descriptive inadequacies.

It motivates depicting a more “global” model of communication (Figure 4), as a realistic alternative to “overt” communication as the outer frame for investigating the ways prosody and prosodic icons work with words.11

![Figure 4. “Global” communication: Presenting – showing – meaning continuum](image)

Such an enlargement of the scope under which communication should be seen and understood, as in Figure 4, has a side interest as well. What is schematized in Figure 4 is human cognition as a whole; at the same time human cognition’s specificity is schematized – and its task, namely: wiring (networking) the diagram’s two extremities, connecting common linguistic meaning (as used in communication) to one’s private sensory-motor experience (see Varela et al. 1993; Núñez 1999; Rohrer 2007).

### 2.3 A third model of communication?

As Sperber and Wilson (1986/1995) showed,12 the Gricean turn achieved a shift from a first model of communication, the “code-model”, to a “cognitive (or inferential) model”. The code model was popularized among linguists by i.a. R. Jakobson’s work; according to this model, communication is seen as a process by which Sender-Speaker encodes thoughts (“content”) in words, and the Receiver-Hearer accesses encoded thoughts by a decoding process – that is “thought duplication” from speaker into hearer.

Grice (1975) demonstrated that much of the content retrieved by the Hearer is not linguistically encoded. Instead, it is inferred through non demonstrative

---

11. Developed in Auchlin (forthcoming).
inferences that take as a premise the assumption that Speaker and Hearer cooperate ("Cooperation Principle"). Relevance Theory’s generalization of Grice’s insight leads to a “Cognitive model of communication”, according to which communication is a modification of the Hearer’s cognitive environment (a modification in the strength with which s/he entertains assumptions) through the decoding of linguistic form and the processing of inferences from an ostensive stimulus.

The third model considers and defines communication as a “co-experiencing” process; it is an enactive and embodied model of communication, according to which communication is the emergence and guiding of the experience of sharing unfolding meaning over time, that is, meaning meaning, as distinct from sharing meant meaning, be it by decoding or inferring.

Let’s examine within this framework how prosodies can directly shape experience.

2.4 Prosodies and experience shaping

Prosodies shape speech experience in time and in audible spectrum dimensions according to distinct mechanisms. In the time dimension, speech rate, duration of syllables and phonemes and relative duration, beat and rhythm elaboration, all impose direct experience modeling. I cannot hear slower or faster than you talk, in ordinary oral communication. In audible spectral energy dispersion, ranging from f0 and f0 variation detection, to more complex processes such as vocalic timbre and formants position perception, it models experience in a less direct way, that relies on active speech perception processes (Skipper et al. 2006), and requires some gradual self-synchronizing to speech in order to be the case. Both dimensions may present singular manifestations that show their relative independence (§§ 2.4.1; 2.4.2). Yet most elaborate iconic formations involve both (§ 4.3).

2.4.1 Speech rate, rhythm and tempo

Let’s see with a simple example how rhythm determines experience.  

(2) Tout est à nous, rien n’est à eux. Tout ce qu’ils ont, ils l’ont volé

13. Following Varela et al. (1993); Núñez (1999), i.a.

14. This is also the test for auditory prosodic analysis: slow down and re-iterate utterances – though keeping in mind the whole shape (Auchlin & Simon 2004).

15. This is not to be confused with natural speech rhythm study – Guaïtella (1999); Barbosa (2006); Auer et al. (1999), i.a.
As a political slogan, the two pairs of sentences are stressed in a binary way, compatible with a marching tempo; but it can also be pronounced with a ternary rhythm, like a waltz – which, in turn, produces a somehow strange result.

Both rhythmic schemas provide cues to interpretation: the first one sounds combative, determinate; the second one displays the words in a non-offensive framework. One may derive very different implicatures from one or the other rhythm.

But the point is that, in both cases, the hearer is prompted to tap out the rhythmic schemes of the speaker’s tempo, according to his own bodily motion. It is in his own beat-tapping that he finds evidence for further interpreting the lyrics. In other words, the hearer must somehow physically instantiate the tempo’s characteristics: this is where information that may – but not necessarily – be followed by various implicatures comes from.

In this example, the march or waltz tempo is not only an un-cancellable premise, it also is inescapable, unavoidable. The tempo is a dynamic motor icon of the march/waltz, that acts as a direct experience driver. Direct experience driving (dynamically shaping) is both prior to linguistic meaning formation (i.a. tempo is given before linguistic constituents’ closure) and (then) independent of it. This distinct layer of information is bound to the utterance’s properties, and it is addressed to the body, and sensory-motor activation, actions and perceptions. We’ll examine further and more sophisticated examples below.

2.4.2 Audible spectrum: Frequencies codes?
The audible spectrum and the sound of the human voice is a mechanism for iconic content manifestation and communication, as has long been noted (Fónagy; Ohala; see also Simon & Prsir this volume).

The sound of the human voice is commonly analyzed in an intonational dimension, concerned with f0 variation, and a higher frequency dimension, that of energy distribution such as in superior formants’ variation, that sums up in vocalic timbre and voice quality. Both determine specific facets of experience.

On the f0 side, noticeable insight was brought by Ohala (1980, 1994), Gussenhoven (2002) with the “f0 code” hypothesis. F0 coding is iconic “by nature”, as it indicates direct (factual or alleged) analogical mapping between f0 properties and the signal producer’s properties.

16. “Everything belongs to us. Nothing belongs to them. All that they have, they stole it.” Motto of the French Nouveau parti anti-capitaliste, April 2008 campaign.

17. “Contextual cues”, according to the C.A. framework, see Gumperz, Auer et al. i.a.; Auchlin et al. (2004) for enactive elaboration of this concept.
F0 code’s basic (hypothesized) ethological function (Ohala 1980) is that it may help prevent unnecessary, fruitless or hopeless conflicts, by letting virtual conflictants evaluate “by ear” their chances to overcome them – or not. This resource is further exploited in speech interpretation (see Footnote 5) – this is not at issue here. What we want to draw attention to is that, in order to be of any use in this basic ethological function, f0 code must enter a two place predicate relation, [BIG (x,y)], and not a one place predicate [BIG (x)]. The first argument (x) is the “sound source”, and the second argument (y) is necessarily “myself”: f0 code’s basic use involves an implicit comparison between the sound source organism’s properties and one’s own properties, as corporeal schema (see Footnote 3). One’s own experience and self-representation is a necessary information source for f0 code to operate – though this might remain implicit and unnoticed.

However, f0 code, per se, doesn’t account for all f0 variation along utterances. F0 variation can also indicate effort change, as coded by the “effort code” (Gussenhoven 2002; see also discussion in Wilson & Wharton 2006:1569), as in the two examples below:

(3) Ce sont eux désormais qui assurent la réputation d’la ville basque

(4) Les TGV Aquitaine arrivent jusqu’en gare de Biarritz

Both examples\(^{18}\) display the same global opposition between two melodic schemas within the utterance, a smooth melodic descent as opposed to a “sawtooth” schema\(^{19}\) – though in reverse position in the two examples. In both cases, smooth

---

\(^{18}\) Drawn from French radio France-Info’s chronicle, same speaker (see Burger & Auchlin 2007; Goldman et al. 2008 for prosodic study of France-Info phonostyle).

\(^{19}\) Prosograms (Mertens 2004) of both utterances are in annexes.
melodic, effortless descent contrasts with an effort-consuming high-low-high display; in both cases the contrast is in phase with the informational status of concerned segments, minimizing effort on the topical part, and maximizing effort on “new”, or relevant parts of the utterance. This may be seen in formulating the questions those utterances seem to be designed to answer, respectively “Who assures Biarritz’s notoriety?” and “Where do TGV Aquitaine trains arrive?” All of this is “effort coded” – a natural signal that undergoes “decoding” procedure by the hearer.

The two examples differ in at least one important respect: whereas in Example (3) melodic display and effort coding is entirely “absorbed” in marking informational status of the two parts, it is not the case in (4).20 In Example (4) first part, effortless melodic descent iconically connects with the part of the utterance representing railroad travel down to Biarritz, hence adding perceptive evidence to it – that is, smooth, effortless and comfortable travel. In the utterance’s temporal unfolding, arrival at relevant information coincides with arrival at travel destination, Gare de Biarritz. Melodic manifestation, in turn, acts as repeated (three times) effort in braking to slow down the moving vehicle, which is another dynamic and motor icon.

This is emergent, second level, “experiential blending”. It is a complex dynamic icon that builds upon two successive iconic formations, both relying on a general frame of moving on a vehicle, inertia, etc. It involves analogical, iconic, transfer from source (utterance) to target (travel) domain; but its emergence crucially depends: (1) on local association with the meaning of utterance parts, and (2) on the utterance’s meaning as a whole. Neither of those conditions is satisfied in Example (3).

F0 coding hardly seems to operate here, or it does so in a complex manner. F0 variation is used as a diagrammatic icon21 of selected properties of a different scale activity that projects them onto perceptive (melodic) evidence.

In the same way, intonation may present manifest iconic potential that do not reach “uptake”, as in following example:

(5) les bi tants ton de Mo dane

20. Though, according to Burger and Auchlin (2007), part of such extra-effort is devoted to building up France-Info’s phonostylistic identity.

The waveform iconic potential would clearly access some uptake with other words, like “un voilier navigant en Méditerranée”. But in the initial case, continuous f0 variation doesn’t enter further elaboration.22

2.4.3 Phonatory posture imitation through formants (proprioceptive formant analyzer) – speech motor imitation

F0 is only part of the energy distribution and discrimination on the audible spectrum. Frequency coding (as the basic ethological component) is also concerned by global energy distribution onto high or low frequencies. As proposed by J. Ohala (1980) global high frequency energy dispersion, as a result of smiling, might as well explain the origin of the smile – that is, acoustically presenting a non-aggressive, if not submissive, attitude. In contrast, “0-mouth” vocal emission produces more low frequency resonance, hence sounding authoritative.

However, this proposal is not concerned by detailed energy distribution, as in formants inter-relations, canonical or non-canonical placement, or voice quality; yet, this level of granularity is required by vowel discrimination, both standard and non-standard. Neither is this proposal concerned by the question of what the receptors of conveyed information are.

The standard (canonical) or non-standard positioning of formants may be seen as a symptom of phonatory posture, or a “natural sign” of that posture (with prior reference to our first language norms). It may in turn further be blended with higher level effect, such as phonostylistic differentiation associated with phonatory posture (Fónagy 1983, i.a.). However, the pathway through which the human hear detects detailed formant placement is not the inferential one they would deserve if they were treated as signs; instead, formant placement detection seems to involve proprioceptive response through phonatory postural imitation.

Could this be seen as a case of “mind-reading” as suggested in Wilson and Wharton (2006:1565), resulting from a “special-purpose inferential mechanism” such as the one activated by gaze direction detection? Probably not: it is not a case of mind-reading, it’s a case of (automated) action imitation.

Recent work on imitation and mirror neuron systems that underlie and determine imitation (Arbib 2006) supports this position. This work brings new evidence for automated action imitation; something that may underlie and support superior interactional functions, whether it be known as synchronization in interaction studies (Auer et al. 1999, i.a.) and ethology, or as empathy in phenomenological psychology (Chertok & Stengers 1989; Cosnier 1994).

22. Except prosodically enacting France-Info’s identity features – which indeed it does.
Skipper et al. (2006)’s findings show that acoustic and visual-gestural information play complementary roles in active speech perception, not just redundant ones. Their observations show that neural implementation is such that visual modality plays a much more important role in triggering pre-motor systems than auditory modality does, in monitored speech perception experiments (such as McGurck-MacDonald effect – below § 3.1.3).

The reported situation is that as far as visual and auditory modalities cooperate in saturating experience, they show dissymmetries as to the degree to which they trigger pre-motor system activation. In mono-modal conditions however things may be somehow different: one might tend to more actively exploit present information in order to gain finer synchronization and action prediction.

How far does phonatory posture and the resulting spectral energy distribution trigger an automatic imitation process is not clear, and would deserve a more in-depth study. Yet, vocalic formant position appears to be a strong imitative trigger. As illustrated during the conference, an artificial elevation of the first or second formant, or both, over an utterance, produces a sound such that automatic phonatory posture imitation occurs. What is accomplished then is an on-line proprioceptive formant analysis.

3. Conceptual blending framework

3.1 Blending

In order to apprehend and schematize general configuration such as it appears in Examples (2) and (4), let us consider here Conceptual Integration Theory’s framework (or blending, in short) (Fauconnier & Turner 2002; Coulson & Oakley 2000, i.a.). Blending is a general theory of how the human mind works and makes sense of actions, perceptions, and conceptual elaborations. It is based on previous work by Fauconnier (1985; 1994) on Mental Spaces that was designed for linguistic semantic description.

Blending, in short, is a process by which information belonging to two (or more) different input spaces are selected and extracted from each, projected

23. Whether, and to what extent, such motor imitation depends on hearer’s knowledge of the spoken language is part of what should be studied.
24. Using Praat (Boersma & Weeninck) formant manipulation, linear shift of automated detection of f1, and f2, ±150Hz. Thanks to J.-Ph. Goldman’s Praat scripting.
25. More precisely, a general framework to understand how the human mind works.
in a “generic space” (which washes them off some – but not all – of their original input spaces’ associated data) and are then “recomposed” through compression into a new mental space, which is a blended space. This framework may be used to describe very different facts, ranging from complex narrative elaboration down to perception: “Integration networks are pervasive in perceptual and conceptual human life” (Fauconnier & Turner 2000: 293):

When we look at the Persian rug in the store and imagine how it would look in our house, we are compressing over two different physical spaces. We leave out conceptually all the actual physical space that separates the real rug from our real house. When we imagine what answer we would give now to a criticism directed at us several years ago, we are compressing over times.

(Fauconnier & Turner 2000: 297)

Simple linguistic examples can be drawn, among others, from compound nouns: interpreting a land yacht requires blending together two distinct spaces associated respectively to land (and ways to travel over it; boats do not travel on land; cars do; …) and to yacht (meant to travel on the sea; big luxury boat; lots of people on board; …). (see Figure 9 in annexes). Counterfactual construction (if I were you…) is another linguistic example – among many others.

3.1.1 Perception
The blending framework also applies to basic cognitive activity such as perception. According to Fauconnier and Turner (2002: 78–80), perception is the blending of cause and effect: “integration of cause and effect is the central feature of perception.” In a less straightforward manner: “(...) the perception of a single entity, such as a cup, is an imaginative feat still very poorly understood by neurobiologists. The perception available to consciousness is the effect of complicated interactions between the brain and its environment. But we integrate this effect with its cause to create emergent meaning: the existence of a cause – namely the cup – that directly presents its effect – namely, it’s unity, color, shape, and weight, and so on”.

(ibid., 70).

3.1.2 Levels specificity
Several authors (Coulson & Oakley 2000: 191–194; Bache 2005) recall that conceptual blending theory has suffered criticisms regarding its ubiquity: in short, “everything is blending” and “blending is everywhere”. In order to face the ubiquity problem, Bache distinguishes three different levels of blending, from simple (and automatic), first level blending, to complex blending processes. Bache argues that in order for the distinct input space’s information to integrate, it must be dis-integrated, extracted from input spaces; dis-integration in this sense is a more
basic operation than the blending (integrating) operation. This distinguishes first level blending, named “binding” (Bache 2005), whose input space information is not open to dis-integration.

Perception is a 1st level blending phenomena. Conscious, deliberate, dis-integration of distinct sources (input spaces) information is not possible in (ordinary) perception; we cannot reverse the binding process in order to isolate input spaces’ (bound) information. We cannot for example mentally dis-integrate saturation, luminosity, environment reflection and contrast, out of a “green spot”. It is a green spot. In Fauconnier and Turner’s terms (3.1.1), the (blended) effect cannot be reverted back to its (input) causes.

Second- and third-order blends, however, “reflect higher-level mental operations on the results of basic first-order compression and integration” (Bache 2005:1622). Those allow for dis-integration of input space information. Second-order blending “refers to what Turner (1996) describes as the integration of basic abstract stories with abstract grammatical structures to produce actual grammatical constructions.” (ibid.). We won’t comment on those levels here.

3.1.3 Mono- and inter-modal perceptual integration: “Stroop-effect” and McGurck-MacDonald effect

Interesting cases for perceptive integration are those implying clashing input space information. Let’s mention two. The first, the “Stroop-effect”,26 shows the limits of possible perceptive integration by a hard-to-blend stimulus; in contrast, the second, the “McGurck-MacDonald effect”, illustrates what kind of (creative) achievement – a strong and robust illusion – may be blended out of partially clashing inputs.

“Stroop-effect” is a well-known case of cognitive disfluency induction. It consists in presenting conflicting information from distinct input spaces: the word green written in red, blue in green, and so on. It is mono-modal as it only relies on visual perception. Mono-modally, it distinguishes alphabetic linguistic input (“decoding”) and color perception. Such are the two input spaces: different information layers. They are forced to coexist, due to perceptive co-occurrence of word-and-color, but due to their respective properties, this is not easily achieved. In other words, the two input spaces are bound by a strong ‘substance of’ relation (or inherence), whose projection and compression to experiential integration requires a strong, voluntarily allocated, extra cognitive effort. The expected result of this extra cognitive effort, its desired output, is in unifying experience as a holistic homogeneous field. This desired output format, notably, acts as well

in constraining input properties – by some “integrability threshold”, as with the Stroop effect, as it does in boot-strapping selected input space properties, as we will show with the McGurck effect.

Testing the Stroop effect should help the reader instantiate both what is called here field of experience, and what kind of experiential output hard blending may cause in that field: either extra cognitive effort in sustaining the trial, or nervous agitation.

The “McGurck-MacDonald effect” (or McGurck effect) is another interesting case of perceptive integration; it implies cross-modal blending. The effect is demonstrated in an 8-second long video, that shows a man’s face pronouncing [ga-ga, ga-ga, ga-ga]; this is what tested subjects extract from the face when attending to the video without sound. At the same time, the audio track says [ba-ba, ba-ba, ba-ba], which, again, is what subjects extract when only hearing. Strikingly enough, in multi-modal integration conditions what is “heard” is the man saying [da-da, da-da, da-da].

This is a strong and constant illusion. In multi-modal (audio-visual) conditions it is hardly possible to experience the same signal either as [gaga] or [baba], yet each of these are in the input; and the global (perceived) signal is only a construct elaborated out of these two diverging inputs, and which is neither.

Linguistic constraints involved in this case are minimal; they are neither conceptual nor representational, but low level and underspecified. It is only the case that someone is speaking (part of Generic Space), not a case of his saying anything. Then no complex top-down bootstrapping illusion might be suspected (as is the case in film doublage, for example, or seemingly in prominence perception, as reported by Goldman & et al. 2010).

![Multi-modal perceptive integration](http://psiexp.ss.uci.edu/research/teachingP140C/demos/McGurk_large.avi)

**Figure 5.** Multi-modal perceptive integration

27. In “Hearing lips and seeing voices”, as McGurck and Mac Donald (1976) put it; demo at [http://psiexp.ss.uci.edu/research/teachingP140C/demos/McGurk_large.avi](http://psiexp.ss.uci.edu/research/teachingP140C/demos/McGurk_large.avi)
The McGurck effect is a somehow special case of binding or first order blending (Bache 2005). Input source information at this first level is integrated in such a way that it cannot encompass the reverse process of dis-integrating from itself; the data are perceptively integrated as objects of a single, unified, apprehension. Compression of both audio and visual input space information is due to overall scene construction at experiential level: we see a speaking person, expecting him to not be able to articulate one sound with his mouth and emit another acoustically. But in the experiment’s setting, information of both input spaces is dis-integrated.

The equivalent to McGurck effect blending, in Stroop effect space, would be, say, to perceive as green the word yellow written in blue – which is not the case.

3.2 Material anchors

Blending can also involve “things”, or material anchors, as noted by Ed. Hutchins (2005). Hutchins’ examples include a wide variety of things that operate as “material anchors” to blends: natural objects (hands, landscape, mountains) as well as artifacts (banknotes; graves; watches, dials, and so on). The latest are themselves built upon basic blending operation, like time compression (let alone the ontological and phenomenal “linearity of time” compressed and projected to a periodic circular movement, for watches). (see Figure 10 annexes for a schematic material anchor blend).

3.2.1 Speaking and writing

Fauconnier and Turner (2002: 210–212) do consider writing and speaking as cases of material anchoring for blends. Blending may be seen in the fact that one may abstract from one medium to the other, from writing to speaking, and vice versa – like in Fauconnier and Turner’s example of the girl reading a letter from her fiancé and “hearing” him, in a blended space where he is close to her and speaks to her. This relies on cross-space mapping, “(…) evolved by culture, for connecting equivalent classes of sounds to equivalent classes of marks (…), a category mark like boy, boy, boy, BOY, boy… to a category of sounds that consists of all the ways of pronouncing the word “boy”.” (ibid., 211). Such cross-space mapping and the use of material anchors for it “depends on a very powerful prior conceptual blend that compresses a certain infinity of marks (boy, boy, boy, BOY, boy…) into a single entity, the written word ‘boy’, and that entity itself is construed as identical to another compressed infinity, the spoken word ‘boy’.” (ibid., 211). The reading activity provides another connection with material anchoring for blends: reading requires “mapping speech in time along linearly ordered locations from left to right horizontally on the page, and understand that at the end of the
line, the speech jumps back to the beginning of the next line, and that turning the page, the most commonest action we take with a book, has no counterpart in the speech space.” (ibid., p. 211).

3.2.2 More material anchoring for speaking and writing
In these cases however, the considered blending operations amount to reducing phonetics to phonemics; in this view, a phonetic, acoustic manifestation is entirely blended to a linguistic manifestation, a phonemic construct. In the same way, the reading activity is considered by reducing the material anchor and its perceptive counterpart to its functional minimum – that is, accessing linguistic messages along the lines. Indeed, this process is at work. But it doesn't give a satisfying account of how phonetic or typographic detail and granularity may act as complex material anchors for blends, nor of what kind of blend emerges. The following paragraph on experiential blending develops this.

4. Experiential blending

4.1 The experiential blending

“Experiential blending” is the integration of abstract conceptual representation (as delivered by linguistic chain treatment, i.a.) and self internal sensory-motor experience of the material properties of utterances (Auchlin 2003). It is regarded as a specific case of blending for two reasons. First it consists in integrating substantially distinct input space information out of distinct levels of blending. One input space is the here-and-now experience of the utterance and its material properties – as perceptively constructed; another input space is linguistic and conceptual elaboration. The other reason is that the blended outcome is experiential: it may consist in experiential enrichment, or embodiment, of concepts or representations (“second level”, see below). More generally, and basically (“first level”, see below), it consists in “human-like” fashioning of speech events (as compared to “machine-like” as actually offered in automatic telephone assistance or other automated verbal interfaces).

Experiential blending is a case of material anchoring; yet it is a special case: here material anchoring is not occasional or episodic; at least for the first level, it is systematic.

4.2 Levels of experiential blending

4.2.1 First level experiential blending
A first level of experiential blending achievement is “structural”, it includes the elaboration of four-dimensionality of internal experience, space and temporality.
It also includes the naturalness of speech signals (or human-like), something that computer speech and speech technology still cannot master – except in copying and assembling natural speech segments.

It is a case of binding, with automatic integration of input space information, and no accessible dis-integration of input space information. Its experiential achievement may to a certain extent be compared to stereophonic audition: right and left channels carry slightly different information; channels differences cannot be perceived as such; instead, perceptive binding outcome is three-dimensional elaboration, and perception of sound in space. Such is the experiential blending of speech: it binds words and clauses meaning to perceptive evidence from utterances, out to speech experience qualities, such as its naturalness, its human likeliness, its actual connection with the experienced world, its depth or superficiality, and so on. However, as binding phenomena, these are phenomenologically transparent: the attention can hardly focus on them. Except when “something goes wrong”.

4.2.2 Second level experiential blending
A second level of experiential blending is emergent, and includes specific rendez-vous between those two input space configurations and their properties, between what is said, and how it’s said.

In his Grammaire temporelle des récits, M. Vuillaume (1990) elaborates out similar cases of integration from “Utterance input space” and “narrative input space”, named respectively “secondary fiction” and “primary fiction”. Vuillaume’s blends are bound to verbalizations like “Au moment où débute ce récit, …”; “Revenons ici quelques mois en arrière”, etc. They establish cross-space mapping, in using reference to “espace parcours” in secondary fiction (deictically; revenons ici quelques mois en arrière; etc.) to determine reference in narrative space, or primary fiction.

Vuillaume’s cases rely on specific cases of verbal token-reflexivity: deictic expressions centered in the writing (or reading) process and connecting to the narration of related events. The utterance space is seen and built up from linguistic input space; what is relevant here is the utterance as type, not as token. Prosodic manifestations, for their part, are bound to utterance as token; their reflexivity is due to the inherence relation they entertain with the linguistic chain. Hence some kind of token reflexivity is at work much more systematically with prosody.

28. This is a ‘qualitative shift’, according to Bateson. Such a shift is alleged for experiential blending.

29. Best example might be the the “lack of naturalness” reproach that former Text-to-Speech concatenation synthesis faced.
A very simple graphic presentation here is preferable to acoustic representation, for illustrating the connection between content and utterance. Let us compare: big to small, and small to big. Such clear-cut contrast probably never occurs naturally; the examples still delimit the (blended) space where it builds up.

4.3 Experiential blending and iconic emergence

4.3.1 “Experiencing budget” blend

Let us examine an example. The text says:

(6) L’inconvénient de nombreux crédits hypothécaires (a) c’est que parfois votre budget est beaucoup trop serré pour pouvoir rembourser facilement (b) et que parfois (c) vous avez les moyens financiers de rembourser davantage (c) de même avec de nombreuses assurances solde restant dû (d) vous payez toujours le même montant par mois (e) comme si vos revenus n’évoluaient jamais (f) Chez Delta Lloyd (g) vous avez des formules flexibles (h) ou flexibles (i) enfin, qui s’adaptent à l’évolution de votre vie (j)

This example displays different successive experiential blends involving prosodic icons, temporal (through speech rate and speech rate variation) and melodic ones, as well as both, temporal and melodic. They all aim at letting the hearer improve the different life conditions associated with evoked abstract concepts, introduced by the initial clause (a): “budget serré”; “budget (trop) large”; “conditions


31. The following examples are taken from commercial advertising. Why? Basically, commercial advertising is onerous; it has to maximize its impact on audience, and this is the business. It then can be seen as a (kind of) natural laboratory developing potentially innovative communicative short devices (clips) – hence handy examples. The acted production and the commercial and media settings are irrelevant regarding the present problematic.

32. The inconvenience of numerous mortgages (a) it is because sometimes your budget is squeezed (tightened) far too much to be able to pay off easily (b) and that sometimes (c) you have the financial means to pay off more (c) also with numerous insurances settles remaining due (d) you always pay the same amount a month (e) as if your income never evolved (f) to Delta Lloyd (g) you have flexible formulae (h) or flexible (i) finally, which adapt themselves to the evolution of your life (j).

33. See (partial) prosogram – Figure 11 annexes.
non adaptables”; or the advantage of advertising Brand, “formule flexible”. Initial clause (a) is pronounced in a “neutral” way – as are the two following “headers” (d) and (g), and the “conclusion” (j). Figure 6 below is a minimal diagrammatic representation of the first three constructions.

Figure 6. Improving budget blend

Three successive prosodic icons emerge from utterances’ temporal and spectral input space information; in association with linguistic input space information they allow an experiential blend to emerge.

Segment (b) presents a first prosodic icon, by the coupling of fast speech rate (9.5 syll. per second) with tense voice and high pitch (mean around 260Hz). Its iconic value is to present tension, stress and lack of time. It is associated to the straight budget concept (from linguistic input), due to temporal contiguity (however at a larger scale than for McGurck effect) and substantial inherence. The hearer’s experience of stressful life caused by too-expensive monthly payments and a low budget is the emergent experiential blended output.

More local experiential blends may arise in this framework. Let us mention one: “budget trop serré” (too tight budget), and the voice with pharyngeal tension, integrate onto an experiential blend of feeling strangulated by the tightness of the
budget. This more local blend relies only on partial linguistic structure, and on part of spectral input space, voice quality, not on f0, nor on temporal input space. However, as hypothesized before, voice quality might be a strong imitational trigger: this should ground the blend in the hearer’s actual experience.

By contrast with (b), segment (c) has a very slow mean speech rate (4.2 s/s), lax voice, low pitch (f0 mean close to the floor, around 100Hz; the difference with the preceding high pitch is more or less 19 semitones). It iconically presents quietness and relaxation. Linguistic input activates “moyens financiers de rembourser davantage” (the financial means to reimburse more) – which is considered here as a disadvantage. This gives rise to a more complex experiential blend. Quietness icon blends, on one side, with partial linguistic structure “vous avez les moyens financiers” – you are rich enough, with the experiential blend of pleasant, cool and comfortable life as output; otherwise, it blends more indirectly with the disadvantageous aspect of the situation. For this contrasted frame to emerge, one has to attach inferences associated to the last adverb, “davantage” (last clause ‘more than you do’), especially regarding financial consequence: paying more means faster closure of credit, which reduces costs. This could re-frame the slow tempo as too slow. But in fact, it does not seem to do so. A probable reason is that a pause before the subsequent segment hardly gives enough time for such a complex elaboration.

The third iconic formation is named “staccato”: syllables of segment (e) are events of the same short duration, more or less 6 syllables per second, clearly separated from each other by short pauses; this isochronicity is associated with mid–high isotonous f0; due to staccato mode, vocalic parts of syllables are short and tense, and the whole signal contains significant parts of consonant articulation – mixing pre-plosive silences and inter-syllabic pauses. This constitutes a dynamic icon of a rapid mechanical hammering. In association with linguistic input information, “vous payez toujours le même montant (…)” (you always pay the same amount), distinct regular payments are projected onto distinct regular syllables, materially anchored in the hearer’s direct experience. In the blended space, the “disadvantage” conceptual contribution is clearly instantiated, through mechanical and unnatural prosody.

A last iconic formation in this example is not represented in Figure 6. Segments (h, i) are about the alleged advantage of proposition (g). In linguistic input space (i) – flexible ou flexible – there is no possible disjunction relation between connected elements. The related terms are intonations on “flexible”: the first one is standard, while second one makes a more than an octave f0 jump between two syllables – hence instantiating in intonation a proper icon of ‘flexibility’.34

---

34. This example shows as well that intonation contours may enter compositional linguistic format as arguments around a connector. This specific integration case is not examined here.
Such L – HL intonation may, in addition, convey some more conventionalized meaning, such as an admiring attitude (as could be expressed by whistling the same tune). This could help instantiate the alleged interest of advertised quality in the blend – giving rise to a (frozen) iconic counterpart of its desirability.

These iconic formations involve Effort coded, as well as f0 coded, information. But it is not clear, however, how much of an emergent effect comes out of any prosodic coded process at all. So far, blended information from prosodic input space only anecdotally, but not systematically, involves those codes.

4.3.2 “Running-talking” experiential blend
Let us examine another example of complex experiential blending – with quite different prosodic iconic input.

The following radio advertisement clip features a male speaker that speaks while running. This is highly manifest, and recognizable: short breath groups; strongly audible in and out breathing; strong beat effect due to running, at about 3.5 steps per second; bird song in the background signals that the man is running outside.35

(7)  
huh huh
   en ce moment huh
   la fiat punto trois portes huh
   est à partir de onze mille cinq cent quatre vingt dix euros seulement huh
   chez votre concessionnaire fiat huh
   à Nantes huh
   moteur huh un litre trois turbo diesel multijet huh
   direction assistée huh huh
   vitres électriques
   radio cd huh fermeture centralisée huh climatisation huh
   une punto trois portes à partir de onze mille cinq cent quatre vingt dix euros
   capital remise déduit huh
   à c’prix-là huh non seulement je cours chez fiat huh mais en plus
   j’arrive le premier

35. huh huh/right now huh/the three-door fiat punto huh/starting at only eleven thousand five hundred ninety Euros huh/at your local Fiat dealer huh/in Nantes huh/one point three huh turbo diesel multijet engine huh/power steering huh huh/huh power windows/Radio/CD player huh huh power door locking huh air conditioning huh huh/A three-door Punto starting at eleven thousand five hundred ninety Euros/Including a discount huh/At that price not only am I running straight to Fiat huh but I’m getting there first.
At the end of the clip, the run receives a verbal and conceptual explanation: the man is running in order to arrive first (J’arrive le premier) at BrandName’s shop, in order to benefit from the announced low price sale. Then, verbal information allows for integrating experiential evidence that, though strongly manifest, wasn’t integrated up until then.

This is, in many respects, a very special case of “Production Code” (Gussenhoven 2002) usage: breathing in and out rapidly, and splitting the verbal stretches accordingly, reflects (as a symptom) the runner’s activity of running, not his activity of speaking and organizing his utterance. It is a natural sign, as defined by Wilson & Wharton (above). Yet it strongly shapes the talk; and it imposes this shape on hearer experience: noisy breathing, level with speech, unary pulsing, and ad hoc text segmentation (breathing and major breaks as in transcription).

In an advertisement setting such as Example (7), the ventilatory symptom of running is acted; it is then shown (as part of the communicative intention), not merely “accidentally” presented (as Example 1’ plural mark). As Wilson and Wharton (2006: 1576) put it, “(…) natural prosodic signs of physical states (…) [breathlessness, i.a.] contribute to speaker’s meaning when ostensively used (…).” It clearly is the case here. What is prosodically shown (as if it was just presented) is an icon of the symptom. However, as in previous cases, it is a dynamic and motor icon; what is extracted from the utterance input space where this icon builds up is its dynamic and motor properties. Its contribution regarding speaker’s meaning is not internal to this meaning; it is in its external, experiential framing. What it contributes to is the embodiment of the concepts and representations activated by linguistic treatment.

In short: the hearer shares the speaker’s run; close to the end, he understands where he is running to; three steps ahead, he learns that the speaker arrives first. The hearer projects himself in a run – hopefully inhibited at motor level, but not at a schematic level (which is a first level of “abstraction”) nor in neural specific activity (Skipper et al. 2006).

36. At second 22 over 25.
37. Note that effective activation of the motor system is the goal of such communication, just as for Example 2. Whether it is obtained or not measures the degree to which advertising or political slogans succeed, in obtaining the desired answer as a “perlocutionary effect”. Activation of dynamic motor icon (in the experiential blend) prompts readiness for action (in the hearer’s world).
38. Be it the only one required for integrating speech segmentation, and noisy breathing out.
Indeed, such projection is part of this ad’s “system” (“experiential program”), it is part of the communicative act: the man is running, then he moves; but his voice does not fade away. So, in listening to the clip, the hearer is moving with him. Indeed, if not, just five or six steps away one would no longer hear the running speaker, or he should speak louder and louder, then shout – which he does not. Thus the listener is, simply by paying attention to the ad’s clip, involved in the experience of running without effective motor activation.

A schematic representation of Example 7 can be drawn as Figure 7:

Figure 7. “Speaking while running” blend

The general schema of this blend is very simple, as it matches utterance space duration and characteristics onto a run to, and arrival at, a car dealership. Speaking while running unfolds the car’s characteristics and the interest of the sale, which
finally appears to motivate the run. The more powerful global experiential achievement is the projection of hearing the ad (utterance input space) onto running with a man to a cars’ dealership (blend space).

This global achievement is anchored in the hearer’s experience by token-reflexive deictic expressions in linguistic input space: “en ce moment” (“right now”) at the beginning, which will frame the whole stretch of talk ahead; and, at the end, “j’arrive le premier” (“I’ll be the first one there”) which, in this context, acts as a performativa, or is true just by being pronounced.

5. Conclusion

To what extent iconicity at large is marginal (Hagège 1982) or pervasive (Fónagy 1999) in natural language, from the lexicon and phonetics to syntax (Van Langendonck 2007) was not at issue here. We did not look at iconicity within language – but within language use in utterances or discourse. Yet it is unclear if language can be studied at all if not in use.

Assuming that, at least some prosodic resources do exhibit iconic properties, this communication addressed two interrelated questions: where do prosodic iconic manifestations combine, mix with the linguistic chain, with content and related inferences? And how is this achieved?

I tried to show that prosodic iconicity, whether temporal, melodic, or both, is a matter of the hearer’s experience of speech. Answers to the where question were sought for within a broad, experiential and embodied framework (following Rohrer, i.a.), which implies re-defining (verbal) communication as building up coordinated speech experiences among participants, or “co-experienciation” – amounting to an alleged third model of communication.

As for how prosodic icons combine with verbal content of utterances, an answer was proposed in using Fauconnier and Turner’s Conceptual Blending Theory framework. The specificity of linguistic and prosodic icons mutual contribution, which takes place in the embodied interpretation process, or the utterance’s experience, close to Hutchins (2005) material anchors for blends, was labeled experiential blending. If prosodic icons are in some way material anchors

41. Or experiencing speech.
for blending linguistic and conceptual input, the emergent and creative part of experiential blending output was claimed to be experiential and not (only) conceptual.

In an experiential perspective, the distinction between percepts and concepts is not as clear-cut as it is in Cartesian linguistics (Lakoff & Johnson 1985). But the difference – be it minimal – between percepts and concepts plays a crucial role in experiential blends with prosodic icons: provided that both linguistic expression and prosodic icon, by virtue of their material contiguity, can be given a common focus, the difference, metaphorically, acts in the same way that the (minimal) difference between left and right audio channels in stereophonic audition works: it is responsible for a qualitative shift, adding a new dimension that neither input could create separately.

Indeed, as for 3-D audition, the emergent reality is in subjective experience, not literally in the signal itself. Ways to objectively study it are as fuzzy as the subjective experience is evident, and somehow unquestionable. In particular, what the conditions and constraints for experiential integration to occur are is far from clear. Further exploration is needed that would bring some refinement within Blending Theory, but also in our conception of prosody and its configurations, including iconic formations, as well as, ultimately, in our conception and knowledge of language, as accessed by means of its usage.

6. Annexes

Figure 8. Prosograms (simplified) for Examples 3 and 4

42. Or not occur, as in the Stroop effect, and Example (5) cases.
Figure 9. The land yacht blend (Fauconnier & Turner 2003: 67)

Figure 10. A conventional conceptual blend (a), and a conceptual blend with a material anchor (b; squares) (Hutchins 2005: 1557)
Figure 11. Prosogram for (part of) Example 6

References


All rights reserved