The Topology of Auditory and Visual Perception, Linguistic Communication, and Interactive Written Discourse

Eli Dresner
Tel Aviv University

Abstract
In the first section of this article, a basic aspect of the spatial topology of auditory perception is contrasted with that of visual perception: Our ear is not sensitive to the direction from which it receives stimulus, while our eye is. This simple observation is shown to be of significance to the communicative characteristics of spoken vs. written language. Then, in the main part of the paper, this theoretical framework is applied to the domain of computer-mediated Interactive Written Discourse (IWD). It is shown that IWD incorporates topological perceptual characteristics of both spoken and written language, and it is argued that these characteristics should not be ignored or given up when synchronous textual CMC is technologically developed and upgraded.

1. Introduction
In recent decades a series of influential thinkers have assigned paramount significance to the emergence of writing and printing, arguing that far-reaching developments—social, economic, religious, cultural, and even cognitive developments—can be traced back (albeit not mono-causally) to these inventions (Innis 1951, McLuhan 1962, Havelock 1963, 1982, Goody 1986, Ong 1982, Eisenstein 1979, Olson 1994). On the other hand, there are those researchers who argue that the claims of the above-mentioned theorists are more often than not unsubstantiated speculations, or even patently false. Thus, for example, Lloyd (1979) raises doubts concerning Havelock’s claims as regards the importance of writing for the emergence of Greek Culture, Scribner and Cole (1981) famously criticize claims to the effect that literacy enhances logical reasoning, and, more

1 I am grateful to the referees of language@internet for their comments. This research was supported by a grant from the Burda Center for Innovative Communications at Ben-Gurion University of the Negev.
recently, Jones (1998, 2002) argues forcefully against the significance ascribed by Eisenstein
(1979) to the technology of print.

On the far speculative end of the manifold of claims (regarding script and print)
under debate stand proclamations to the effect that the mere visuality of written language (as
opposed to the auditory nature of speech) has far reaching cognitive and communicative
ramifications. McLuhan’s (1962) talk of change in perceptual balance received much
criticism, even ridicule, and Ong’s (1982) elaboration of this line of thought falls short of
providing a sound theoretical basis for it. Even Olson (1994), who is (and presents himself
as being) ‘in the same camp’ with these thinkers, says that although this avenue of inquiry
may ‘ultimately be relevant’ (p. 38), so far it consists in a mere metaphor, not an
explanation.

The objective of the first section of this paper is to take the first steps towards
answering Olson’s challenge—i.e., to show how the two different perceptual modalities
support different communicative uses of language. This will not be done through an appeal
to a mysterious balance-change in human perception. Rather, note will be taken of some
basic physiological differences between our visual and auditory perceptual mechanisms, and
it will be argued that these differences are of substantial communicative significance. Then,
in the second and main part of the paper, the suggested theoretical framework will be used
to provide an analysis of Interactive Written Discourse—i.e., computer-mediated, textual
synchronous interaction. It will be shown that IWD incorporates topological perceptual
characteristics of both spoken and written language, and that it can therefore be said to
occupy a middle ground between spoken and asynchronous written language in a very basic,
perceptual sense. This observation, in turn, will be claimed to be of significance to the future
development of IWD in particular, and of CMC in general: Computer mediated
communication allows us more ‘perceptual freedom’ than any other medium (i.e., it allows
us more freedom in using different perceptual modalities in conveying information), and
therefore making the best use of CMC requires that we understand the perceptual trade-offs and balances involved.

2. Spatial Topology of Auditory and Visual Perception and Linguistic Communication

2.1 The Physiology of the Eye

The following description of the structure and function of the eye (Kolb and Wishaw, 2001) is partial and brief. Its objective is to highlight only those points which will be of relevance later in the discussion.

The stimulus of our visual system is light—electromagnetic waves of very short wavelengths (400-700 nanometers). Light rays enter the eye through a small hole—the pupil. Then they pass through the lens, which bends them to a greater or lesser degree, and continue traveling through the eyeball’s interior until they reach the retina. The retina contains a dense system of photoreceptors—cells that convert the electromagnetic stimulus into neural electric pulses (emitted by attached neurons). The nerves from all parts of the retina are bundled together into the optic nerve, through which the neural information is transferred to several areas of the brain where it is processed.

Note that light enters the eye only through a very small hole because this way every area in the retina is accessible to light reaching the eye from a specific direction. (If the eye had a wide entrance, then each part of the retina would be hit by light from a wide angle of directions.) This way the two-dimensional visual field is recreated (albeit upside down) as a two-dimensional picture on the retina. This picture, in turn, is transferred ‘intact’ backwards into the brain, in the following sense: The neurons coming from different parts of the retina are ‘kept apart’ (although they are bundled together into the optic nerve), and thus each neuron is known to convey stimulus from a given direction.

Basic visual stimulus, then, as perceived by each eye, is of a two-dimensional spatial structure, representing the spatial structure of the visual field in front of the eye at each
moment. (That is, spatial relations among the objects that we see are reflected by spatial relations among the areas in the retina that are stimulated by light coming from these objects.) As a result of this basic stimulus, coming from both eyes, we are able to create a three-dimensional visual representation of the world around us.

2.2 The Physiology of the Ear
The anatomical structure and functioning of the ear are extremely complex; their essentials were understood in full only in the second half of the 20th century, and are much less a part of common knowledge than their visual counterparts (Kolb and Wishaw, 2001).

The outer extreme of the ear is the pinna; it deflects the sound waves into the ear canal, where they travel until they reach the eardrum and move it in a certain frequency. The eardrum, in turn, passes the vibrations through a series of three tiny bones (the ossicles) to the cochlea—a snail-like shell which is filled with liquid. Inside this liquid is found the basilar membrane; this membrane bends due to vibrations induced in the cochlea’s liquid by the ossicles. The basilar membrane is covered with hair-cells, which convert the movement of the membrane into pulses in the neurons attached to them. Different sound frequencies cause bend-peaks in different locations in the membrane, which, in turn, cause different neurons to fire signals. The nerves from all locations in the basilar membrane are bundled together into the auditory nerve, which carries the signals into the brain for processing.

The basilar membrane of the auditory system can be said to be similar in function to the retina in the eye—both are the loci where stimulus is translated into neural activity. Moreover, both sub-organs (the basilar membrane and the retina) are structured so that different areas on their surface are sensitive to different stimuli, and thus their spatial structure helps represent the structure of the stimulus at any given moment and transmit it into the brain. However, on the background of this similarity there should be noted the following great dissimilarity between the two organs. As we saw, the spatial structure of the
retina represents a structure that is spatial as well—i.e., the visual field. The outer part of the eye is constructed so as to enable such representation, by allowing light from a given direction to reach only restricted parts of the retina. The basilar membrane, on the other hand, functions differently—different areas in the membrane are not sensitive to sounds from different directions, but rather to sounds of different frequencies. And of course, the outer part of ear is structured in tune with this functionality: The pinna and eardrum do not differentiate between sounds coming from different directions, but rather collapse all waves reaching the ear into a single (albeit complex) vibration pattern that is passed inwards.

This difference between the ear and the eye is of course not arbitrary; it can be said to originate (in the Darwinian sense) from the great difference in wavelength between sound waves and (visible) light waves. The greater length of the former makes the sound waves that reach our ear a mishmash of waves from a variety of sources, waves that got around obstacles and diffracted from all kinds of surfaces. These waves cannot be perceived by the ear as coming from a given distinct direction. (Analogously, radio receivers pick up radio waves without being pointed in this or that direction.) The infinitesimal wavelength of visible light, on the other hand, allows us to think of rays of this light as consisting of little particles—photons—that definitely do reach the eye from a distinct direction, and are perceived as such.

The observation as regards the structure of the ear that is key to our purposes here is this, then. The basic auditory stimulus, as perceived by each ear, is indeed structured (single-dimensionally), but not as a representation of any spatial structure of sound sources. As far as the single, ‘brain-independent’ ear is concerned, sound does not reach the ear from a two dimensional matrix of differentiable sources (as is the case with the eye), but rather from a single source, a single point—‘out there’. Now of course, the data from both ears is put together and correlated with other sensory modules, and thus enables us to hear things as coming from this or that direction, from near or from afar. (This is on a par with the way the
two-dimensional pictures produced by each eye are put together into the three-dimensional visual grasp that we have of the world around us.) However, these achievements, allowed by elaborate processes, do not undermine the characterization of the ear as a perceptual mechanism made above (and the matching, different characterization of the eye). As will be argued below, this feature and its visual counterpart are of significant communicative importance.

2.3 Speech and the Topology of Auditory Perception

As noted in the introduction, Ong (1982) is one of the major thinkers who ascribe importance to the difference in perceptual modality between spoken and written language. The main characteristic of sound in this context, according to his view, is its evanescence (p. 32): “Sound exists only when it is going out of existence. It is not simply perishable but essentially evanescent, and it is sensed as evanescent.” Now this basic distinction of Ong’s is important, and Ong draws substantial consequences from it. However, the evanescence of the spoken word is not the only significant property that arises from its auditory nature. Another such property is related to the topological aspect of auditory perception discussed in the previous subsection.²

We noted above that the ear senses the outside world as a single ‘out there’—i.e., it does not discriminate among sounds coming from different directions. For this reason, we can hear sounds coming from a given direction without paying auditory attention to it. As an example, consider the (much worn out) scenario of someone shouting ‘fire!’ in a concert

² Let it be clear that neither Ong’s analysis nor the one presented below is claimed here to exhaust the discussion of the differences between spoken and textual discourses, and the various grounds for these differences (among which are technological and perceptual grounds, like those discussed here, but also social and cultural ones, that arise in specific contexts). Our focus here on a single line of argument is methodological, and should not be mistaken for a deterministic, mono-causal outlook. For example, it is certainly not denied here that differences in the production (rather than perception) of written vs. oral language have significant ramifications, or that the use of intonation and visual cues are key characteristics of standard, face-to-face auditory conversation.
hall. The shout is effective because it can be heard by the many people (most probably all of them) who were listening to the music, and thus did not give any prior perceptual attention to the person who shouted. Admittedly, they may hear the shout as coming from a certain direction (or distance), but their ears do not have to be directed in that direction. In comparison, imagine how much less effective it would be for the person issuing the warning to raise a written sign with same message: only those looking in his direction (and not at the stage) would receive the message.

As yet another example, think of the much more common situation of the phone ringing at someone’s home. In order for her to hear the ringing, the person being called does not have to focus her auditory attention on the phone—she can listen to music, or be engaged in conversation. When the phone rings its sound will be added to the single, topologically point-like summation of sounds that is ‘in front’ of the person’s ears, so to speak, and if the sound is loud enough it will be acknowledged and acted upon. Again, if the phone had only flashed a light when activated (as some phones do, on top of the sound they emit) it could have been perceived only if the addressee’s gaze happened to be on it (or near it) at the time.

Let us see now how this general characteristic of hearing bears upon the subject matter of our interest here, namely auditory conversation. In typical conversation situations, every participant in the conversation can hear every other participant (as well as himself). Therefore, according to the above discussion, everyone does hear everyone, in the following sense. It is not the case that a person’s ears are pointed in this rather than that direction (or at this rather than that speaker); as we saw, the ear cannot be spatially directed. Rather, for each and every one of the participants in a conversation the auditory contributions of all participants (including himself) are placed in the same, single basket—the summation of sounds that is gathered by his ears at any given moment. Consequently, because there is no significant perspectival difference among conversers, all of them are auditorily exposed to
the same sound-basket. Using a visual metaphor, the situation can be summarized thus: all conversers have their auditory gaze fixed on the same buffer, on which they all place their messages and get (i.e., hear) everyone else’s messages (as well as their own).

Let me first provide some evidence supporting this claim, and then defend it against an important reservation (or misunderstanding). The natural source of the required empirical (and conceptual) support is the field of Conversation Analysis (CA)—the domain of study which has the working of auditory conversation as its primary concern. If we take a look at the methodology, concepts and major findings in this field we can see that they all support (or presuppose) the above observations concerning auditory perception. Here are three examples.

First, consider the data used by many conversation analysts (Psathas, 1995; ten Have, 1999): transcribed conversations that are viewed (and presented) as single buffers in which messages accumulate through time, and which are available to all participants in the conversation. A single auditory conversation is not presented from different perspectives: it is assumed that the same transcription describes what all participants hear. This assumption, which echoes our observations here, is key to the objectives of researchers in CA (Psathas, 1999, p. 2): they are interested in regularities and structures that help shape the interaction, and that are therefore available to the participants themselves. Hence the fact that the researcher looks at the conversation from a single perspective implies that she assumes all participants share this perspective, i.e., that they have similar auditory inputs. Admittedly, several of the formative studies in CA were made on telephone conversations, where this assumption is more plausible than in other conversation situations (Sacks 1992, Schegloff 1968, 1979). However, it is clear that some of the findings of these studies are supposed to have wider applicability, and hence so is the assumption underlying them.

Second, some of the most prominent concepts in CA presuppose the topological structure described here. Consider the basic notion of a conversation turn. Turns, by
definition, take place one at a time, and are perceptually available to all players. Therefore they presuppose a single place (a game-board, or a buffer) in which each player makes his move, and which is perceived by all—exactly the topological structure of auditory perception that we have been talking about here. Alternatively, consider the notion adjacency pair (Schegloff 1968): In what sense are two messages that make up such a pair adjacent? Clearly not in space, but also not simply in time. (For example, when they do not together form a conversation thread they do not constitute an adjacency pair.) Rather, the messages come one right after the other in the single auditory buffer that is available to all participants in the conversation. Thus these important concepts presuppose (and therefore support) our observations here.

Finally, concrete research in CA supports these observations. As an example, consider the renowned Sacks, Schegloff and Jefferson (1974) study of turn-taking: both the data and the resultant theory of this research presuppose the structure articulated here. As of the data, the fact that there is little overlap and minimal gaps between turns (p. 700) shows that there is indeed a single, collectively-attended auditory conversational buffer that is kept non-empty and not overcrowded most of the time. And the theory that is proposed to explain this fact (and others) implicitly presupposes the same accessibility pattern. For example, in clause (1.b) of the theory it is stipulated that (p. 704) “first starter acquires rights to a turn….” Such a rule cannot make sense unless all participants receive the same auditory input, and can therefore agree (albeit unconsciously) who started first.

We see, then, that there is ample evidence supporting the claim that the structure ascribed above to auditory perception (based on the ear’s physiology) plays an important role in conversation.³

³ It should be acknowledged that the examples considered here are concerned with specific kinds of spoken interaction (typically a conversation between (two or more) equals), and are situated in a specific (i.e., western) cultural context. Thus further inquiry is required to ascertain whether and how these examples, and the analysis
It could be objected at this point that the above-presented picture of face-to-face conversation is misguided. As many studies have shown us (Goffman 1981, Goodwin 1981, 1994), the spatial relations among the participants in a conversation do matter, there are many visual cues (e.g., gaze) that are highly important for normal conversation, and hearers’ attention is directed at different people (as well as other objects) in different stages in the conversation. Conversation is a perceptually, socially and culturally complex phenomenon, and therefore it is patently wrong to describe it as a situation in which participants receive similar perceptual inputs, and in which spatial orientation is of little significance.

The answer to this objection is that the observations concerning conversation it invokes are of course true, but that nevertheless it rests on confusion. In order to see this, we need to distinguish between conversation as a whole on the one hand, and the bare auditory experience of participants in conversation on the other hand. (Put differently, the distinction is between listening and hearing, where listening is construed as taking part in auditory conversation, and hearing as receiving auditory perceptual input.) Conversation itself is indeed a highly complex and varied phenomenon, and as such it would be clearly absurd to ascribe to it the topological structure formulated above. However, the auditory input stream of conversation-participants is another matter; it is just one of the factors (or components) that together constitute conversation. The above topological characterization is claimed to apply only to this factor, and to help explain how this factor contributes to the whole conversational experience.

As a concrete example of this point, consider the interplay between voice and gaze, as discussed, e.g., in Goodwin’s (1979) analysis of the production of a single sentence during face-to-face (dinner table) conversation. Goodwin suggests the rule that (p. 106)
“[w]hen the speaker gazes at a recipient he should make eye contact with that recipient,” and says that the speaker can request such eye contact *(ibid.)* “by producing a phrasal break, such as a restart or a pause, in his utterance.” Now one reason for which such a phrasal break is useful for calling the hearer’s attention (more than a facial movement, for example) is that the speaker’s voice is available to the (would-be) hearer without his having to direct his ears to the speaker—exactly the topological point made above (and exemplified also in the fire warning example). Thus, it is not claimed here that gaze and other perspectival factors do not play an important role in conversation; they obviously do. Rather, it is only argued that the non-spatial aspects of hearing play an equally important one.

2.4 Writing and the Topology of Visual Perception

The property of the written and printed word (as opposed to the spoken one) that receives much attention from many writers is that it is an *object*—a time-surviving, movable and reproducible object. This focus of attention is stated explicitly by Ong (1982, p. 31), but it is implicit in various other writings as well. For example, Innis’ (1951) discussion of time- and space-biases of (written) communication media is concerned with variations in this property. As another example, consider Eisenstein’s (1979) argument to the effect that the printing revolution made a significant contribution to the emergence of modern science. This argument is based primarily on the observation that printing allowed the production of many identical copies of (old and new) texts, thus enabling (i) the same text to be available to many individuals, and (ii) the same individual to have available to him many texts. Here again the focus is on the printed text being an easily reproducible object.

Now the characterization of the written/printed word as an object clearly does not exhaust the communication-relevant features that distinguish written from spoken language. The discussion in the foregoing sections of this paper suggests another such feature (albeit among many others): because the text in front of us has the persistent existence of an object
(albeit sometimes a short lived object), and because we perceive this object visually, we experience text as spatially structured. The words, sentences and paragraphs on the page (or screen) are located on a two-dimensional surface, and have spatial relations among them: they are above or below each other, to the left or to the right of each other, one word is between two others, etc. Moreover, these written linguistic particles have metric relations among them as well: the notions of distance (i.e., far and near) apply to them. Now as elaborated in section 2.1, our vision allows us to grasp these spatial and metric relations: the image produced by each of our eyes at any given moment is spatial and metric in nature, mirroring the structure of the objectively-existing text.

Spoken language, on the other hand, is not perceived spatially. In section 2.2 we noted the physiological background for this fact, and in the previous section (2.3) we focused our attention on its communicative advantages—i.e., how it enables auditory communicators to be ‘on the same page’. Here we note another aspect of this same fact: literally speaking, we do not experience spoken words as above or below each other, nor to the left or to the right of one another. One word (or sentence) might be louder than another, or either lower or higher in pitch, but this is not spatiality in the basic sense we are considering here. (By way of clarification, consider the words in sign language: they are evanescent like their spoken counterparts, but are still perceived as being located (albeit not concomitantly) in space and have spatial relations. Spoken, auditory words are different in this sense—they are not spatially located.)

Note that the above stated distinction is consistent with our experiencing both written and spoken language through time, with respect to which we do have spatial intuitions (i.e., we think of time as a one dimensional space and consistently apply metric conceptions to it). Indeed, we do experience spoken words as being produced sequentially through time,

---

4 As elaborated below, this is not to deny that any reading of a given text is also temporally structured.
temporally preceding or succeeding one another, or being temporally distant or close to each other. (We have a similar experience of written words in any specific reading—we perceive them in a temporal sequence.) However, these temporal relations are distinct from the spatial relations among linguistic particles, which can be found in written text and not in spoken language. The spatial structure of text represents the temporal sequential structure noted above, and (possibly) much more.

A key feature distinguishing writing from speech, then, is its spatial perceptual nature. Now as in the foregoing analysis of auditory perception and speech, it should be asked whether there is any communicative significance to this characteristic of visual perception of written and printed language. Ong (1958, 1982) answers this question positively: Our capacity to take in the logical and conceptual complexity of (some kinds of) written text is dependent on our ability to perceive it as a space, in which, for example, we can move back and forth and on which a metric is imposed – i.e., distances among points. Conceptual relations among ideas and thoughts are partially indicated by (and remembered according to) the order in which they appear in the text, as well as by such mechanisms as paragraph, clause and chapter structure – all of which are visually represented. The spatial grounding of syntactic and semantic linguistic structure is more evident in the case of print than in handwritten text, but it is operative in the latter case as well (Goody 1986).

Let it be acknowledged again (as was done in the introduction) that early deterministic claims connecting medium and conceptual complexity are almost commonly agreed today to be misguided. Tannen (1982, 1989), Chafe (1982) and Biber (1988), for

---

5 Let it be acknowledged here (on the same vein of remarks already made in subsection 2.3 above) that reading should be distinguished from the bare seeing of text: the first phenomenon is set in a social, cultural and historical context while the second is not. No deterministic connection is stipulated here between the perceptual characteristics of seeing text and the communicative and cognitive ramifications of reading; rather, the perceptual aspects of seeing are claimed to have some influence on the higher level phenomenon of reading, an influence that it is valuable (by my lights) to investigate.
example, argue convincingly that in our day and age there is no complete match between the
linguistic expression mode (spoken vs. written language) and discourse features: there is
speech with many of the characteristics typically ascribed to written language and vice versa. (Indeed, Tannen (1982) holds that there is a continuum between the oral and literate
types of discourse.) However, as noted by Olson (1994), these observations do not stand in
contradiction to the claim that writing, when introduced into a culture, helps put in place the
far ends of this (possibly multi-dimensional) continuum, and that in contemporary western
culture, at least, its ends are often associated with the two (written and spoken) modes of
linguistic expression.

3. Auditory and Visual Perception and Interactive Written Discourse
We now turn to apply the conceptual framework presented above to the domain of IWD—
Interactive Written Discourse. (The term is due to Ferrara, Brunner and Whittemore (1991)).
IWD can take many forms, and has been around now for several decades. However, in this
discussion we shall consider one of its most prevalent contemporary incarnations, namely
Internet textual chat. The most popular technology of Internet chat is IRC (the Internet Relay
Chat protocol), but there is a huge host of Web-based chat services with basically the same
familiar format. The chatroom consists of a data-base, which a (large) number of
participants can be logged on to at any given moment. The contents of the data-base are
presented on the computer screen of each participant, and are pushed upwards as new
messages appear. Each participant can write a message at any time, and once she presses the
‘enter’ key on her computer the message will be added to the data base and appear
immediately (or possibly after some lag) on the screens of all participants. There are many
additions to/variations on this basic format—e.g., the use of icons of various degrees of
graphic sophistication, options to switch to private interaction between participants, etc.—
but these need not concern us here.
3.1 Auditory-like Perceptual Structure of Textual Chat

Interactive Written discourse is usually construed as a communicative phenomenon that is somewhere in between spoken and written language—the very term ‘Internet chat’ indicates this conception. But why is written language of this particular form thought of as similar to spoken discourse? The most natural and common answer to this question is that IWD is synchronous and interactive, like spoken conversation. From this basic analogy are claimed to follow many other affinities between textual and ordinary conversation—e.g., the contextuality of the contents expressed, or the fact that messages are unedited. (It is impossible to edit auditory conversation turns once they have been produced, and impractical to do much editing of messages in a lively textual chat.)

Analyses on these lines are correct, but I argue that they are missing an important part of the picture. There is another kind of affinity between auditory and textual chat, on the basic level of perceptual topology. This affinity can be noted on the basis of the considerations made in the previous section.

We noted that the participants in a given conversation are (by and large) receiving similar raw auditory input, as the ear does not enable spatial differentiation. The single, spatially-unstructured unit that everyone’s auditory intake consists of can be thought of as a communication buffer—a buffer which is perceptually available to all participants all the time, and which (auditory) messages are placed on and taken from. As noted above, a visual analogue to this auditory situation would be a screen available to all participants in the conversation all the time, which (visual) messages are placed on and taken from. Once this analogy is made, however, the beans are spilled: Standard textual Internet chat consists in such a visual analogue to the auditory set-up. In a simple (i.e., single-window—see section 3.2 below) chatroom situation all participants sit in front of their computer screens. All of them are seeing the same thing—the text lines accumulating in front of them. As opposed to visual perception in spoken conversation, where each participant sees a completely different
picture, in textual conversation vision functions somewhat like hearing in auditory discourse—it enables mutual focus on the buffer on which communication takes place.

We see that the affinity between ordinary and textual chat goes beyond (or, rather, deeper) than synchronicity. The structure of mutual visual perceptual intake in computer mediated textual chat is topologically similar to its auditory counterpart. This analogy in perceptual structure between the two cases may contribute to an explanation of the social characteristics of IWD—in particular, its appeal as a medium that brings people together. In section 2.3 it was noted that the topology of auditory conversation enables participants to be ‘on the same page’ perceptually, and thus possibly helps them to be together socially as well. Similarly, IWD enables chat-participants to be ‘on the same page’ literally, a fact which may contribute to their feeling of being together socially (Turkle 1995, Danet et al. 1997, Herring 1999, 2001).\footnote{An interesting twist of this idea concerns the textual representation of physical action in IWD, as discussed, e.g., in Herring (2001, p. 632). In group face-to-face interaction a given action performed by one of the participants is often not noticed by all others, because people may be looking in other directions. (This is the directedness of vision, discussed in subsection 2.2.) The textual actions in IWD, on the other hand, are possibly more commonly observed, because of the auditory-like focus of visual attention in chat, as described in this section. Thus linguistically represented action can be more commonly shared than its real counterpart.}

This observation echoes, in a way, Anderson’s (1983) claims with respect to the contribution of printed newspapers to the evolution of nationality. Anderson argues that the fact that large audiences were exposed to the same contents at (approximately) the same time, and, moreover, that they knew they were so exposed, helped form in them the conception that they were all members of a single, national body. Similarly, it is argued here that the mere basic perceptual topology of computer mediated chat, which allows participants to see the same things (things linguistic, that is) at the same time (and know that they do)—this basic topology may be of social and psychological significance.
Once the structural similarity between spoken conversation and Internet chat is noted (and its significance acknowledged), it is natural to ask whether this similarity is the result of a conscious choice. Was Internet chat designed to have this perceptual affinity to auditory conversation? The answer to this question is negative: The perceptual analogy noted in this paper is not the result of planning. Indications that this is indeed the case are several. First, no mention of the considerations raised here is made in accounts of the development of, e.g., IRC. (See, for example, Oikarinen’s account in http://www.irc.org/history_docs/jarkko.html.) Second, the structure of IRC and its Web-based analogues is evidently the result not of considerations of the kind presented here, but rather of technological context and limitations. IRC has its origins in pre-graphic Human-Computer interface (HCI), where the appearance of text lines one after the other on the (undivided) screen was the default—for HCI, and thus also for CMC. Finally, with the advent of graphic interfaces for synchronous communication the structural similarity we are concerned with here is unthinkingly given up—another indication that the designers and users of the technology are not aware of it. (More on this in subsection 3.3 below.)

We see, then, that the feature of contemporary IWD under consideration here was not a result of planning or design. If this feature is indeed of communicative significance and value (contributing to the success of this new medium), then we have here another example of the intricate relations between technology development and use: useful features are sometimes (some would say often) not the result of design, but rather unintended consequences.

3.2 Visual Spatiality and Textual Chat
We saw, then, that textual chat manifests an inter-personal perceptual topology that is analogous to that of auditory perception in spoken conversation. This fact contributes to the status of this communication medium as standing between spoken and written language. It is
natural to ask now a complementary question, concerned with the discussion of spatial
visuality and linguistic communication in the previous section: Is this discussion relevant in
any way at all to Internet Chat? Is the spatial structure of visual perception operative in
textual synchronous conversation, and is its contribution similar to those (claimed to be)
found in standard, asynchronous writing?

As noted in many places (Ferrara, Brunner and Whittemore, 1991; Werry, 1996;
Herring, 1999), written discourse in chatrooms is usually made up of relatively short
messages. The ideas expressed and the semantic relations among them are often not
intricate, to say the least, and the whole coherence of the communicative interaction is low:
dialogues are not maintained for a long time, and topics arise and are then discarded quickly.
Some researchers attribute these phenomena mainly to features of the medium, while others
put more stress on social and cultural factors.

These brief observations are sufficient to establish the following point: The syntactic
and semantic effects of spatiality on language found in traditional, asynchronous writing are
not prevalent in interactive written discourse, at least as typically practiced today (Herring
2001, p. 617):

One medium variable, however, does exercise a powerful influence over structural
complexity: synchronicity. Just as the structure of unplanned speech reflects
cognitive constraints on real time language encoding, for example in length of
information units, lexical density and degree of syntactic integration, so too

---

7 Typical chatroom messages are shorter than messages in asynchronous computer mediated textual media. I do
not know of any research comparing the length of chatroom messages to that of turns in auditory conversation.

8 For example, Werry (1996, p. 53) notes the relatively long time it takes to express oneself in writing online,
leading to commensurate shortening of the messages exchanged in order to retain conversation flow. Also, he
says (ibid.) that the large number of people taking active part in chatroom conversations creates competition
for the attention of potential listeners (i.e., text viewers), and therefore motivates linguistic expression that is
easy to digest, and hence relatively short and quite often provocative. Stromer-Galley and Martinson (2004), on
the other hand, note the difference in message length and speed of topic decay between entertainment and
political chats; this difference indicates that indeed factors other than technical help shape the form of IWD
interaction.
synchronous modes of CMD [Computer Mediated Discourse – e.d.] impose temporal constraints on users that result in a reduction of linguistic complexity relative to asynchronous modes. It would seem, then, as if the spatial structural resources of written language, which non-interactive textual communication appeals to, are left untapped in such interactive communication (due to its synchronicity, as well as other factors). However, a more careful and thorough examination of contemporary IWD shows this conclusion to be unjustified. The structural features of spatially-perceived language are indeed used in chatroom conversation, but in a manner different from their application in asynchronous writing. Here is how.

As already noted above, and elaborated by Herring (1999, 2001), chatroom discourse exhibits reduced interactional coherence. When the conversation includes more than two or three active participants, the written sequence of messages, which unfolds on the participants’ screens, is a jumble of different discourse threads, unanswered approaches, and system messages. As opposed to the norms of auditory conversation, messages that directly follow each other on the screen are usually unrelated; messages that do form a meaningful conversation thread (among two or more people) are dispersed over the whole linear sequence. Now as noted by Herring (1999), the reasons for this lack of coherence are mainly (1) that the messages are displayed by the system in the temporal order they are received, without consideration of any semantic or communicative factors, and (2) the complete absence of any audio-visual cues that in standard auditory conversation help discourse management so much. Herring points out several ways in which chatroom participants try to overcome this chaotic state of affairs, e.g., by writing explicitly whom they are addressing in their message—Werry’s (1996) addressivity—thus enabling the addressee to ‘pick out’ the message directed to him. Also, she argues that, in fact, this lack of coherence serves some communicative purposes: for example, it makes the whole chatroom scene more light-
hearted and playful. This is why, Herring says, Net chat can be so incoherent yet so popular and successful as a communication medium.

Herring’s analysis, as just described, is insightful and true to the facts. Nevertheless, I argue that the picture she presents of the phenomenon in question is one-sided; in a sense, she looks only at the empty half of the glass. Herring takes for granted the coherence norms of standard, auditory conversation, and (rightly) notes that chatroom written conversation does not satisfy these norms. However, the same situation can (and should) be described also differently: chatroom written conversation has interactional standards of its own, standards that are not (and cannot be) met by auditory, everyday conversation. What are these standards? An elaborate answer to this question is beyond the scope of this paper, but in what follows we discuss one of them, that is relevant to our interests here.

As just noted, in many chatroom conversations there is a concomitant intermingling of several voices and conversation threads, all taking part in the same linear text stream. With some practice, participants develop an ability to differentiate among the different threads within this stream; they can follow and contribute to one or more of these threads (Herring, 1999). Thus, some IWD chatrooms resemble a stereotypical cocktail party (or a dinner table), in which several conversations take place concomitantly, but in which skillful participants can follow and take part in more than one of the conversations. In an auditory situation, on the other hand, one can admittedly catch his name in a conversation going on in another part of the room, but the rule is that we do not, and cannot, follow more than one conversation line for a substantial period of time.

_________________________

9 Of course, there are many IWD environments (and, in particular, many IRC channels) in which the interaction is of an entirely different pattern. As elaborated below, no causal connection is claimed here to exist between medium and conversational structure; the argument is only that the new medium enables new kinds of such structure.
We see that the innovative medium – interactive written communication – leads to another innovation, defined not in technological terms, but rather in terms of the communication pattern: IWD enables a conversation situation where all participants are continually perceptually aware of more than one conversation line. These two distinct (although related) novelties of computer-mediated communication are usually not sufficiently distinguished from one another. Once they are distinguished, however, we may inquire about the relationship between the two: Why does interactive written discourse enable such ‘multi-focal’ conversation?

As in the case of asynchronous writing, considered above, there are several features of the written text that give rise to its communicative characteristics in the synchronous case as well. And as in the asynchronous case, again, one of these features is the durability of the textual word—albeit a short-lived durability, on the computer screen; indeed, Herring (1999) herself notes that one factor that helps chatroom participants stay on top of the conversation is the fact the messages stay on screen for a short while, and thus can be associated with each other more easily. However, in this case too (as in the asynchronous case) the visual spatiality of the synchronous text is operative as well: the structured, metric spatiality of the text helps us keep track of various conversation threads at the same time. We see a sufficiently large segment of the conversation before us at all times, and associate spatial visual relations among messages with relevant semantic relations among them—typically, being components of the same conversation thread. Pictorial processing abilities seem to help us sort out the entanglements of conversation lines.

This claim is supported by the common practice of CMC-users to open several chat-windows together, and follow all of them at the same time (Turkle, 1995). Turkle’s informants report the ease with which this division of attention is accomplished, and clearly one important reason for this ease is the simple fact that the different conversations (or
textual narratives) are located in different areas of the screen. Thus, the visual ability of the user to take in a spatially complex picture at any given moment helps him note immediately which window is active, and to sort out the different happenings in the different chatrooms in which he is participating.

The foregoing discussion can be summarized as follows: The spatial structure of written language consists in a resource that may be utilized in various ways, helping us deal with various kinds of communicative complexity. In some cases of non-interactive linguistic communication, such linguistic complexity is typically longer sentences and lexia, as well as more involved logical structure. For (some kinds of) interactive discourse the complexity absorbed by the spatial structure of text is of a different nature, namely that of various interwoven conversation threads. In both cases, the communicative complexity in question is not caused by spatial structure; rather, it is driven by extra-technological motivations, and allowed by the textual medium. On the other hand, in a weak sense the textual medium can be said to give rise to the communicative patterns that are supported by it—without the use of text these patterns would not arise. However, there is no paradox here—just a typical application of the notion of resource: a resource is such that it is used for reasons that are extraneous to it, yet its availability is also a reason for its use. I believe that thinking in these terms is useful—for the consideration of communication technologies in general, and the case at hand in particular.

In a similar way, we can describe the perceptual unity of auditory perception as a resource that is used in various ways in spoken conversation (as elaborated in subsection 10.2).

This hypothesis is empirically supported in Dresner and Barak (2005), where two textual conversation threads were shown to be more comprehensible when presented in separate windows than when intermingled in the same window.
Similarly, the analogous perceptual unity of textual chat (subsection 3.1) consists in such a resource as well. Thus, we see that textual chat can offer both resources (or features): structured spatiality and unified focus. Of course, both resources cannot be tapped fully at the same time: when extensive use of spatial structure is made, unity of perceptual focus is reduced, and vice versa. However, the advantage of the new synchronous medium is that it incorporates both dimensions, and can therefore give rise to an array of variants combining these dimensions in different ways and to different degrees. (See subsection 3.3 below.)

The value of applying the theoretical framework presented in section 2 to interactive written discourse is twofold, then. Theoretically, this framework enables us to locate IWD vis-à-vis spoken and written discourse in a substantive, informative way. Such a concrete location can hopefully enhance our ability to apply more of what we know about speaking/listening and reading/writing to the study of IWD in particular, and to CMC in general. At the more practical level, the foregoing analysis provides a framework with which to characterize various forms of synchronous textual interaction, assess their advantages and disadvantages (only from the perspectives considered here, of course), and possibly devise new and improved interfaces. I pursue this second direction in the following, final subsection.

3.3 Interactive Written Discourse and Graphic User Interfaces
At present, graphic user interface (GUI) in computer-mediated written conversation is typically applied to introduce extra-linguistic aspects into the communicative interaction; in particular, it is commonly used to accord a pictorial embodiment to the ‘voices’ that take part in the (written) conversation, rendering such conversation closer to ordinary spoken discourse. For example, in 2D graphical virtual reality environments each voice is visually associated (e.g., through cartoon-like balloons) with a picture representing the person speaking, as we associate real voices with people around us. (A good example is the popular
Palace environment (www.thepalace.com).) As computing and data transfer speeds increase, these pictures (often called *avatars*) become more elaborate and dynamic and may even mimic human behavior.

This kind of graphic IWD clearly aims at increasing the similarity between computer mediated written conversation and the spoken variety. For example, efforts at analyzing human speech-related body language and its positive effects on conversation and comprehension have been underway for some time now (Vilhjalmsson, 1997), the results of which may be applied in the development and constant improvement of avatars that simulate human activity relevant to the printed text. The goal of this intriguing and significant research is to render graphic chatrooms a more natural – and consequently a more efficient and effective – communications medium. (Obviously, the development of audio-visual computer-mediated chat, in which the participants can see and hear each other, strives for the same objective, i.e., assimilation of computer chat and ordinary conversation.)

Nevertheless, it is important to note that by increasing its similarity to standard, spoken conversation environments, the graphic chatroom might be relinquishing the distinguishing features of written interactive discourse, some of which were shown above to be desirable. Only one side of the resulting tradeoff is usually acknowledged: By improving emulation of spoken language in computer mediated IWD, we may indeed benefit from various well known and desirable aspects of spoken language that could be gradually introduced; however, at the same time, we may be giving up advantages of contemporary IWD—advantages that may not appear natural to us at present but may be highly beneficial nonetheless. Of course, this principle applies to the development of communication technologies in general (Hollan and Stornetta, 1992): It is not always the case that the best possible emulation of non-mediated (or just familiar) communication avenues brings about optimal communicative results. (For example, switching from auditory communication to
audio-visual communication (which is closer to non-mediated, face-to-face interaction) does not necessarily enhance understanding (Short, Williams and Christie, 1976).)

As a concrete example of the above-mentioned tradeoff, consider the speech-balloon interface mentioned above. If every new message by a certain speaker replaces the old one in this balloon, the spatial continuity that characterizes non-graphic chat, as noted in the previous section, will disappear. Conversation threads will be much more difficult to trace backwards and forwards and consequently harder to follow in the absence of this novel (and for some purposes highly useful) characteristic of contemporary, non-graphic written discourse. Also, speech-balloon interfaces lose the auditory-like perceptual focus of basic, textual chat: When ‘linguistic happenings’ occur in different locations on the screen different participants may have their visual attention focused on different locations, and thus the perceptual unity of the conversation is lessened.

This turn of technological events is not inevitable: Graphic chatroom interfaces need not necessarily emulate standard, spoken conversation. Instead, the same graphic technology can be used to enhance and develop the unique possibilities and advantages of written conversation, even at the cost of greater dissimilarity between it and its “natural,” spoken ancestor. Here are three rough examples.

1. **Branching conversation.** Branching textual dialogue is typically found in asynchronous CMC—for example, in computer-mediated bulletin boards; it is usually not thought of in the context of interactive written discourse. Why is this possible variety of IWD neglected? One reason could be that in auditory conversation there is no coherent branching option. A second reason is technological: as observed in the previous section, non-graphic, character-based IWD, which until recently has been the only kind of computerized written discourse available, consists of a continuous, single stream of characters (even if divided into lines on the computer screen).
The present study and advances in IWD suggests that branching IWD is an option that should be pursued, both conceptually and experimentally. (See (Smith, Cadiz and Burkhalter, 2000), (Viegas and Donath, 1999) and (Vronay, Smith and Drucker, 1999) for discussions of systems that implement this general outlook.) For example, consider a scenario in which one participant in a conversation wishes to comment on a statement made earlier (from which the conversation has since shifted). In spoken conversation, one would have to divert everyone’s attention to that previous statement, thereby breaking the current thread. In a textual framework, however, one could comment on a previous remark and refer to it by some onscreen graphic device on the participants' computer screens. Those participating in (or following) the conversation could then apply their visually-related linguistic capacities to consider the suggested discourse branching while allowing the main thread to continue. Each participant may then decide whether or not to develop the branch, possibly at the expense of what is now the principal line of discourse.

Is this option viable, and, if so, in which contexts? Only further investigation and technological experimentation can tell. However, in terms of the perceptual balance presented in this paper we can say that this (hypothetical) kind of textual chat has at least some ‘theoretical’ merit to it: the auditory-like perceptual unity of ordinary chat is given up here for a concrete conversational gain, rather than for the mere emulation of embodied conversation (as is the case with avatar-and-balloon chats).

2. Multi-Focal Conversation. As noted in the previous subsection, it is common for IWD users to have several open windows on their screen, each presenting a distinct synchronous interaction. In view of the foregoing discussion, it is natural to ask whether a single group of users (two, or more) could find use in leading a synchronous interaction that takes place in several windows at the same time. For example, think of two people who are doing business and gossiping. In auditory conversation, these two activities must be intermingled with other, or else one has to be postponed after the other. In synchronous
textual conversation, however, there is a third option—each activity can take place in a distinct window: this way the two activities are temporally intermingled, but spatially distinct. Now it is reasonable to suppose that if this scenario is viable there may be various connections between the distinct windows in which the (same, multi-faceted) conversation takes place, and thus there will be need of various graphical devices to indicate such connections.\footnote{This suggestion (as well as the previous one) is in the spirit of Nelson’s (2004) vision that in creating digital text we should not attempt to mimic printed text, but rather free ourselves from the latter’s limitations.}

3. Auditory-like Focus. The foregoing two examples were concerned with the spatial character of IWD, and the way it can be put to communicative use. As indicated towards the end of the previous subsection, a complementary resource of IWD is its auditory-like perceptual unity, which also needs to be acknowledged and used. As an extant example, consider again the Palace (\url{www.thepalace.com}) graphic virtual reality environment. An interesting feature of the interface is that messages appear both in balloons (coming out of the avatars’ mouths’) \textit{and} in an ordinary, textual chat window that is located next to the graphic virtual reality window. What is the motivation for this duplicity? Why do users find it useful? A reasonable hypothesis (supported by interviews with several participants) is that the non-graphic chat window helps users stay on top of the conversation, because it accumulates the conversation in a single buffer.\footnote{Similar considerations may be relevant to the design of other graphic environments, and even to audio-visual ones. For example, in the audio-visual popular Paltalk environment there is a standard chat window, and it may be of interest to investigate what its communicative and social functions are. Cursory observations indicate that it does not serve only to enable auditory (or audio-visual) interaction.} Thus the spatial visual characteristics of synchronous text—viz. the fact that the same text can be presented more than once, at the same time—helps support its auditory-like perceptual focus. Are there other ways in which these factors can be fruitfully combined? As in the previous cases, only further theoretical and experimental work will tell.
4. Conclusion

In the first section of this paper it was shown that visual and auditory perception manifest distinct topological characteristics, and that these characteristics are of significance to linguistic communication. Then, in the second part of the paper, it was argued that Internet textual chat possesses perceptual characteristics of both written and oral language, and is thus indeed placed ‘in between’ these two linguistic modes in a perceptual sense (as well as in other senses, which were not addressed here). This observation, in turn, was claimed to have practical importance to future development of textual CMC interfaces: different interfaces exhibit different balances between the topologies in question, and these balances should be acknowledged, analyzed and matched to the communicative purposes that are desired in each specific case.
References


Submitted: 27.02.2005
Review results sent out: 15.04.2005
Resubmitted: 01.05.2005
Accepted: 01.05.2005

Any party may pass on this Work by electronic means and make it available for download under the terms and conditions of the Digital Peer Publishing Licence (DPPL). The text of the licence may be accessed and retrieved via Internet at http://www.dipp.nrw.de/.