Examining the neural mechanisms involved in the affective and pragmatic coding of prosody

Didier Grandjean & Klaus R. Scherer

Swiss Centre of Affective Sciences
University of Geneva
Didier.Grandjean@pse.unige.ch

Abstract

The vocal expression of humans includes expressions of emotions, such as anger or happiness, and pragmatic intonations, such as interrogative or affirmative, embedded within the language. These two types of prosody are differently affected by the so-called push and pull effects. Push effects, influenced by psychophysiological activities, strongly affect emotional prosody, whereas pull effects, influenced by cultural rules of expression, predominantly affect intonation or pragmatic prosody, even though both processes influence all prosodic production. Two empirical studies are described that exemplify the possibilities of dissociating emotional and linguistic prosody decoding at the neurological level. The first study was conducted to investigate the impairments in prosody recognition related to left or right temporo-parietal brain-damaged patients. The second study used electroencephalography in healthy participants to investigate the timing of information processing during emotional and linguistic prosody recognition tasks. The results highlight the importance of considering not only the distinction of different types of prosody, but also the relevance of the task realized by the participants to better understand information processes related to human vocal expression at the suprasegmental level.

1. Introduction

Prosody or intonation is often considered a prime carrier of affective information, a function that has often been neglected in research done in this area (with a few notable exceptions, e.g., [1], [2], and [3]). One of the reasons for this neglect might be the lack of a consensual definition of "intonational form" (see, for example, [4], [5]), which specifies how intonation should be empirically measured. Most work on prosody has been informed by linguistic models of sentence intonation that focus on accent structure and which are based on widely differing theoretical assumptions [6], [7], and [8]. Another factor that has hindered progress in the understanding of emotional communication via prosody is the difficulty of describing the coding principles, that is, the transfer functions, from psychobiological emotion signatures to acoustic patterns in speech.

Scherer and collaborators [3] suggested two general principles underlying the coding of emotional information in speech, *covariation* and *configuration*. The covariation principle assumes a continuous, but not necessarily linear, relationship between some aspect of the emotional response and a particular acoustic variable. Thus, if F0 is directly related to physiological arousal, F0 should be higher in rage as compared with mild irritation. Thus, Ladd and collaborators [2] suggested that the F0 range shows a covariance relationship with attitudinal and affective information in that a larger range communicates more intense emotional meaning. The authors used resynthesis to systematically manipulate the F0 range and variability to investigate the effects on emotion inference. Of all variables studied, the F0 range had the most powerful effect on judgments. A narrow F0 range was seen as a sign of sadness or of absence of specific speaker attitudes. A wide F0 range was consistently judged as expressing high arousal, producing attributions of strong negative emotions such as annoyance or anger. In contrast, almost all linguistic descriptions assume that intonation involves a number of categorical distinctions, analogous to contrasts between segmental phonemes or between grammatical categories. In consequence, the configuration principle implies that the specific meaning conveyed by an utterance is actively inferred by the listener from the total prosodic configuration, such as "falling intonation contour", and the linguistic choices in the context.

The configuration principle seems to determine the coding of pragmatic features of speech, for example, emphasis, or message types such as declarative or interrogative mode. How, then, can this principle code emotional information? For example, final pitch movements are coded by the configuration principle. Final-rise versus final-fall patterns of F0 in themselves do not carry emotional meanings; they are, rather, linked to sentence modes such as question versus nonquestion. However, it can be shown that context, such as type of sentence, affects interpretation. Whereas a falling intonation contour is judged as neutral in a WH-question, it is judged as aggressive or challenging in a yes/no question [3]. Thus, the configuration principle seems to allow coding of emotional content through a deviation from standard linguistic prosody patterns. In perception studies using resynthesis, prosodic configurations such as "uptrend" versus "downtrend" contours, accent height in sentence-final (second) accent position, relative height of subsequent local maxima in the F0 contour, durations of accented syllable, and so forth were manipulated and showed effects on listener judgments [9], and [10].

When will each of these principles be used, respectively? Ladd and collaborators [2] suggested that overall F0 range and voice quality might reflect arousal, whereas differences of prosodic contour type signal differences of more cognitively based speaker attitudes. Alternatively, it could be hypothesized that continuous variables are linked to push effects (externalization of internal states), whereas configurations of category variables are more likely to be linked to pull effects (specific normative models for affect signals or display; see [11]). One way to approach this question is to look at the phylogenetic origin of the coding principles.

In terms of origin and evolutionary development, it seems plausible to suggest that the covariation principle is evolutionarily continuous with the biopsychological mechanism that underlies affect vocalizations in a large number of species. This possibility is described, for example, by the motivational-structural rules suggested by Morton [12] in an attempt to systematize the role of fundamental frequency, energy, and quality (texture) of vocalization for the signaling of aggressive anger and fear, see [13].

In contrast, the configuration principle might be assumed to be an evolutionarily more recent development, based on the emergence of language with its specific design features, including intonation patterns. Affective meaning could be produced by nonstandard usage of these respective codes, depending on the degree of context-dependent emotional marking, see also [2]. If this assumption is correct, one could imagine that the neural mechanisms that underlie the perceptual processing of the two types of affect messaging via prosodic variation are different, both with respect to the neural structures and circuits involved and to the nature and timing of the respective processes.

A preliminary step in the empirical testing of this prediction is an examination of the difference in neural auditory processing of speech samples communicating either emotional content (joy, anger, sadness) or linguistic-pragmatic meaning categories (e.g., statements or questions). If the prosodic communication of emotional content via the configuration principle uses a nonstandard, or marked, version of linguistic-pragmatic prosody identifying message type, it would be useful to first identify the potential neural processing differences between covariation-based emotion prosody patterns and linguistic-pragmatically coded prosodic message types. In this contribution, we present two empirical studies performed in our laboratory that have a direct bearing on these issues with respect to the perception of these two utterance types.

2. Effect of neural damage on processing

The implication of the left (LH) and right hemispheres (RH) in the decoding of emotional and linguistic-pragmatic prosody has been extensively studied through brain-damaged patients. The cue-dependent hypothesis suggests that the LH is specialized to decode temporal information, whereas the RH is involved in processing of spectral cues [14], see also [15]. The involvement of each hemisphere for prosody processing would also be influenced by the task and the significance of the stimulus for the listener during a specific situation (e.g., [16], [17]). Finally, the focus of the participant's attention is also a key question to the understanding of the automatic and controlled processes and the contribution of different lateralized parts of the brain to process emotional and linguistic prosody information [18], [19].

In order to investigate the differential contribution of the LH and RH in linguistic and emotional prosody decoding, Siegwart tested 14 patients with temporo-parietal lesions, 8 left-hemispheric-damaged (LHD) and 6 right-hemispheric-damaged patients (RHD; [20]). The performance of these patients was compared with the performance of 14 controls matched for age, gender, and sociocultural level. The French stimuli ("Alors tu acceptes cette affaire" and "Vous restez à la maison") were produced by two actors (one female), with happiness, anger, and sadness emotional prosody as well as declarative, interrogative and injunctive linguistic prosody. The actors produced the target sentences with either the emotional or the linguistic prosody types alone (simple stimuli), or with a combination of both (composite stimuli)The participants had to perform a semantic matching task, a recognition task, and a discriminative prosodic task with filtered (low-pass filtering) and unfiltered simple and composite stimuli (including both prosody types). The stimuli used to test the ability of brain-damaged patients to recognize the different prosody types were selected on the basis of recognition rates of healthy controls.

The results showed a significant impairment of the performance of LHD patients for the recognition of composite stimuli compared with the controls' performance ($F(1,7) = 10$, $p<.02$; see Figure 1). Moreover, the LHD patients showed a disability related to linguistic prosody during the semantic task compared with the matched controls ($t(7) = 4.11$, $p<.01$; see Figure 2

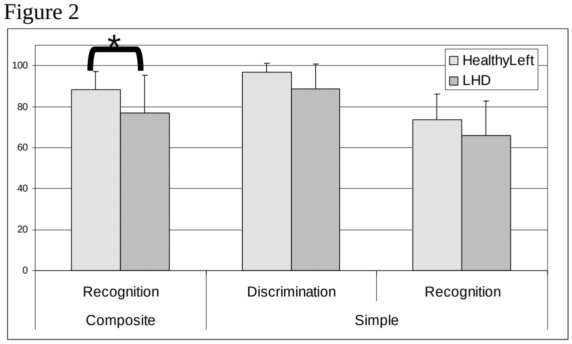

Figure 1: *Percentages of correct recognition and discrimination tasks for composite and simple stimuli for the LHD and healthy matched controls (* = p<.05).*

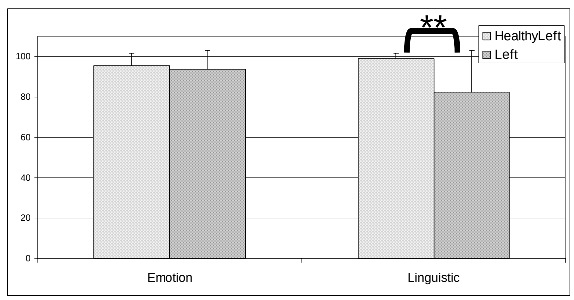

Figure 2: *Percentages of correct responses for the semantic task for emotional and linguistic prosody stimuli for the LHD and healthy matched controls (** = p<.01).*

Note that the LHD patients showed a significant correlation (0.72, $p<.05$) between the performance of a semantic intonation battery and the recognition of linguistic prosody stimuli.

The RHD patients did not show the semantic effect presented by the LHD patients. However, the RHD patients showed an impairment for linguistic and emotional prosody recognition and discrimination for simple stimuli compared with matched controls (respectively $F(1,5) = 16.3$, $p<.05$ and $F(1,5) = 21.1$, $p<.01$; see Figure 3). Together, these results challenge the traditional view of a stronger involvement of the RH to decode emotional prosody compared with the LH. The relative contribution of the two hemispheres in decoding emotional and linguistic prosody features are probably related to subprocesses implicating different parts of the two hemispheres in several information processing steps, see also [15].

These results highlight the importance of considering not only the type of stimuli (linguistic and emotional), but also the task in which the participants or the patients are involved (semantic or prosodic recognition or discrimination tasks, for instance), the task being also related to attentional processes. Moreover, in future studies, the *interaction* effects between linguistic and emotional levels should be addressed more systematically.

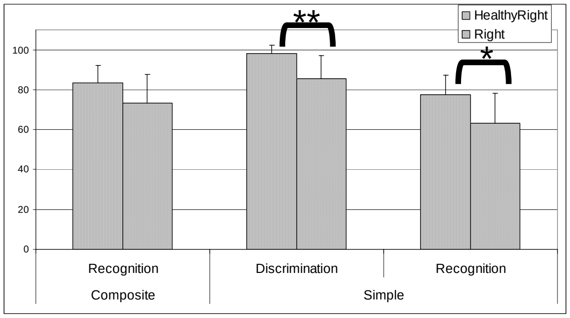

Figure 3: *Percentages of correct responses for recognition and discrimination tasks for simple and composite stimuli for the RHD and healthy matched controls.*

To further investigate these two different processes (linguistic and emotional prosody identification), we conducted an electroencephalography (EEG) study, described below, with healthy participants without brain damage.

3. The dynamics of electrical brain activity in prosody processing

In order to investigate the timing of linguistic and emotional prosody decoding, and the underlying differences in information processing at the central nervous system level, we recorded the electrical brain signals using EEG in 15 healthy participants. We used a NeuroScan system with QuickCap (64 channels) and computed the event-related potentials (ERPs) for each participant and each experimental condition. Three simple French words were used ("ballon," "talon," and "vallon"), with the F0 contour being systematically manipulated using Mbrola synthesis ([21], see Figure 4) to produce happiness, sadness, and neutral emotion expressions, as well as affirmative and interrogative utterance types.

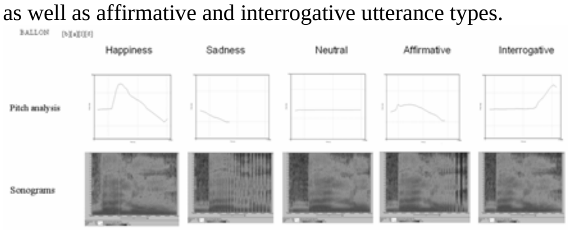

Figure 4: *Examples of pitch analyses and sonograms for the French utterance "ballon" for the different experimental conditions used in the EEG study.*

During EEG recording, the participants had to identify emotional and linguistic prosody, as well as phonemic differences within three different counterbalanced blocks. Time slots for occurrences of specific topographical brain maps obtained by cluster analyses [22] as total average ERPs are different for the three recognition tasks. The results highlight specific processes related to emotional and semantic prosody identification compared with phonemic identification.

Specifically, the three first ERP electrical brain maps (C1, C2, and C3 maps in Figure 5) are common for the different experimental conditions. Between 250 and 300 ms to ~ 400 ms, specific processes occurred for emotional prosodic identification (E map with a right anterior positivity) and semantic linguistic identification (S map with a centro-posterior negativity), demonstrating the involvement of different underlying neural networks subserving these different mental processes. In fact, the statistical analyses show specificity of the maps for both the emotional prosody and the linguistic-pragmatic conditions when compared with the two other conditions, respectively [23].

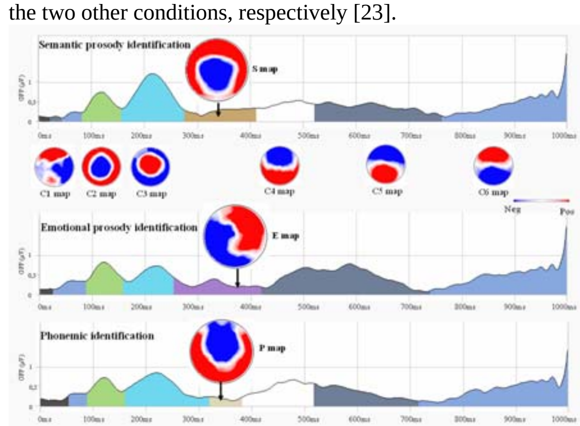

Figure 5: *Occurrence of the topographical brain maps over time (1000 ms after the onset of the stimulus) for the three experimental conditions. The maps are represented on the global field power.*

These results indicate that specific neural circuits are involved in the recognition of emotional prosody compared with linguistic and phonemic identification tasks (P map in Figure 5). A right anterior positivity was measured on the scalp for the identification of emotional prosody; this result is compatible with a previous fMRI study demonstrating anterior activations in right dorsolateral and orbitofrontal regions during emotional identification compared with phonetic identifications of the same stimuli [24]. The involvement of the left part of the frontal region was highlighted in another fMRI study when the participants had to identify linguistic compared with emotional prosodic information [25]. However, the two temporal parts of the hemispheres are differentially involved in different subprocesses that contribute to the recognition of the emotional content of a word or a sentence (see [16]). For instance, different brain networks process temporal information compared with spectral information, respectively, in the left temporal versus the right temporal parts of the brain [26].

The specific electrical map related to the recognition of emotional prosody in this EEG experiment cannot be explained solely by the fact that the intonation contour was modified, because we also used different F0 contours for the linguistic-pragmatic condition (interrogation and affirmative contours). Moreover, the same stimuli were used in the phonemic identification condition, demonstrating that this specific emotional prosody map is not related to the differences of basic acoustical features, but rather is related to the type of the participant's recognition task. This study underlines the possibility of using speech synthesis to systematically modify acoustic features of emotional prosody, inducing different types of categorization processes related to the participant's tasks. In the future, this type of paradigm could allow researchers interested in the understanding of perception of emotional prosody to study the integration of different subprocesses contributing to the subjective perception of intonation in emotional processes.

Further studies are needed to systematically manipulate the different acoustical dimensions involved in different functions at the prosodic level with vocal synthesis. In contrast to fMRI techniques, EEG methods allow the study of not only the interactions of different brain areas in prosodic perception, but also the *timing* of these processes to identify the brain structures involved in prosody perception.

4. Conclusions

The two empirical studies described above emphasize the importance of considering not only the different types of prosody, but also the different tasks and significance of the stimuli or event for the individual. Further studies should systematically address the different contributions of these two factors, as well as the influences of push and pull effects related to covariation and configuration decoding processes.

5. References

[1] Fonagy I., 1981, A semiotic approach to the study of prosodic irregularities, In *Phonologica 1980*, Akten der vierten internationalen Phonologie-Tagung, W.U. Dressler, O.E. Pfeifer, J.R. Rennison (eds). Innsbruck, 137-152.

[2] Ladd, D.R.; Silverman, K.; Tolkmitt, F.; Bergmann, G.; Scherer, K., 1985. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker affect. *Journal of the Acoustical Society of America,* 78, 435-444.

[3] Scherer, K.R.; Ladd, D.R.; Silverman, K., 1984. Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America, 76*, 1346-1356.

[4] Beckman, M.E., 1996. The parsing of prosody. *Language and Cognitive Processes,* 11, 17-67.

[5] Möbius B., 1995. Components of a Quantitative Model of German Intonation. *Proceedings of the 13th International Congress of Phonetic Sciences (Stockholm)*, 2, 108-115.

[6] Bruce, G.; Garding, E., 1978. A prosodic typology for Swedish dialects. In *Nordic prosody,* E. Garding, G. Bruce, R. Bannert (eds). Gleerup, 219–228..

[7] Bruce, G., 1982. Textual aspects of prosody in Swedish. *Phonetica,* 39, 274-287.

[8] Ladd, D.R., 1983. Peak features and overall slope. In A. Cutler and D. R. Ladd (eds). *Prosody: Models and measurements*. Berlin: Springer, 39-52.

[9] Goldbeck, T.P.; Standke, R.; Scherer, K.R., 1988. Techniken der digitalen Signalverarbeitung in der vokalen Kommunikationsforschung. *Psychologische Rundschau,* 39, 191- 200.

[10] Ladd, D.R.; Scherer, K.R.; Silverman, K., 1986. An integrated approach to studying intonation and attitude. In *Intonation and discourse,* C. Johns-Lewis (ed.). London/Sydney: Croom Helm.

[11] Scherer, K.R., 2004. Feelings integrate the central representation of appraisal-driven response organization in emotion. In A.S.R. Manstead, N.H. Frijda, A.H. Fischer (eds). *Feelings and Emotions: The Amsterdam Symposium*. Cambridge: Cambridge University Press, 136-157.

[12] Morton, E.S., 1982. Grading, discreteness, redundancy, and motivation-structural rules. In D. E. Kroodsma; E. H. Miller; H. Ouellet (eds). *Acoustic communication in birds*. New York: Academic Press, 182-212.

[13] Grandjean, D.; Baenziger, T.; Scherer, K.R., in press. Intonation as an interface between language and affect. *Progress in Brain Research.*

[14] Van Lancker, D.; Sidtis, J.J., 1992. The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research***,** 35, 963–970.

[15] Schirmer, A.; Kotz, S.A., 2006. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences,* 10(1), 24-30.

[16] Van Lancker, D., 1980. Cerebral lateralization of pitch cues in the linguistic signal. *Papers in Linguistics: International Journal of Human Communication,* 13(2), 201–277.

[17] Mitchell, R.L.; Elliott, R.; Barry, M.; Cruttenden, A.; Woodruff P.W., 2003. The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia,* 41(10), 1410-21.

[18] Grandjean, D.; Sander, D.; Pourtois, G.; Schwartz, S.; Seghier, M.; Scherer, K.R.; Vuilleumier, P., 2005. The Voices of Wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience,* 8(2), 145-146.

[19] Sander, D.; Grandjean, D.; Pourtois, G.; Schwartz, S.; Seghier, M.; Scherer, K.R.; Vuilleumier, P., 2005. Emotion and Attention Interactions in Social Cognition: Brain Regions Involved in Processing Anger Prosody. *NeuroImage,* 28(4), 848-858.

[20] Siegwart, H.M., 1993. Traitement des fonctions expressive (émotionnelle) et linguistique de la prosodie : une étude de neuropsychologie expérimentale. Thèse no 199. Genève. Suisse.

[21] Dutoit, T., Pagel, V., Pierret, N., Bataille, F., Van Der Vrecken, O., 1996. The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes. *Proceedings ICSLP'96, Philadelphia, vol. 3,* 1393-1396.

[22] Michel, C.M.; Murray, M.M.; Lantz, G.; Gonzalez, S.; Spinelli, L. ; Grave de Peralta, R., 2004. EEG source imaging. *Clinical Neurophysiology,* 115(10), 2195-2222.

[23] Grandjean, D.; Ducommun, C.; Bernard, P.-J.; Scherer, K.R., 2002. *Comparison of cerebral activation patterns in identifying affective prosody, semantic prosody, and phoneme differences*. International Organization of Psychophysiology, August 2002, Montreal, Canada.

[24] Wildgruber, D.; Riecker, A.; Hertrich, I.; Erb, M.; Grodd, W.; Ethofer, T.; Ackermann, H., 2005. Identification of emotional intonation evaluated by fMRI. *NeuroImage,* 15, 1233-41.

[25] Wildgruber, D.; Hertrich, I.; Riecker, A.; Erb, M.; Anders, S.; Grodd, W.; Ackermann, H., 2004. Distinct Frontal Regions Subserve Evaluation of Linguistic and Emotional Aspects of Speech Intonation. *Cerebral Cortex*, 14, 1384-1389.

[26] Zatorre, R. J.; Belin, P., 2001. Spectral and Temporal Processing in Human Auditory Cortex. *Cerebral Cortex,* 11, 946-953.