Keynote lecture

Functions and mechanisms in linguistic research — Lessons from speech prosody

Yi Xu

Department of Speech, Hearing and Phonetic Sciences, Division of Psychology and Language Sciences, University College London, UK

Abstract

Human speech, as a collection of complex phenomena, can be explored from many different angles, and interesting data can be generated with various approaches. However, if we aim to substantially improve our understanding of speech, it is advantageous to focus on the communicative functions and the mechanisms that enable these functions. This paper uses prosodic phenomena in speech as examples to illustrate the necessity as well as potential benefits of such function- and mechanism-oriented approaches.

Key words: communicative function, mechanism, speech prosody

Introduction

Human speech, as a collection of complex phenomena, can be studied from many different angles. This is true not only across broad divisions such as syntax, semantics, phonology, etc., but also within each division. In speech prosody, for example, the same set of phenomena can be examined by asking rather different questions: What is the best way to describe melodic events — in terms of level tone or holistic contour (Bolinger, 1986; Cruttenden, 1997; Ladd, 2008; Pierrehumbert, 1980)? How are prosodic patterns related to syntax (Beckman, 1996; Selkirk, 1984)? What are the prosodic correlates of prominence and boundary tones (Hermes & Rump, 1994; Kochanski et al., 2005; Terken, 1991)? What is the essence of speech rhythm (Loukina et al., 2011; Ramu et al., 1999)? Is compression or truncation the chief strategy of tonal variation in a language (Grabe, 1998)? Should tonal alignment be part of the basic description of a language (Ladd et al., 1999; Kohler, 2005)? While these questions all seem very interesting, there is another question that is rarely asked, namely, is answering questions like these the best way to improve our understanding of speech in general and speech prosody in particular?

What most of these questions have in common is that they are mainly about the form of prosody. That is, they address issues that are directly related to the directly observable properties of prosody, while the communicative meanings associated with these properties are left vague. A good example is the Autosegmental-Metrical (AM) model of intonation, in which the intonation components like pitch accent, phrase tone and boundary

tone are all defined in form, while their functional definitions are relatively vague (Ladd, 2008; Pierrehumbert, 1980), and this is despite deliberate efforts to identify the possible functional meanings associated with the proposed categories (Pierrehumbert & Hirschberg, 1990).

An alternative strategy that has been steadily gaining recognition is the function-oriented approaches or functional approaches for short (Hirst, 2005; Kohler, 2005; Xu, 2005). Nevertheless, non-functional approaches still dominate prosody research. Likewise, in many other areas of linguistics, non-functional approaches also dominate. In the rest of this paper, a case will be made for approaches that focus directly on communicative functions and mechanisms that enable their realization.

The case for functional approaches

Of the many reasons for taking functional approaches seriously, the most fundamental is based on consideration of the nature of speech. This can be illustrated by a comparison between speaking and singing, both of which use the human vocal apparatus to control the melody of vocalization. In singing, every note needs to be as accurate as ±1 semitone from the designate frequency (Dalla Bella et al., 2007; Pfordresher et al., 2010), otherwise the singer will sound out of tune. This is the case because music is form-driven (Patel, 2008), and so perfection in form is of utmost importance. In speech, in contrast, even in a tone language where every tone needs to be quite accurate, the precise F_0 value of any tone varies extensively from speaker to speaker, and from context to context, even among speakers of the same gender and age. But the tones still sound natural and function perfectly, because within individual and each context, different tones are sufficiently dissimilar to each other so that listeners have little difficulty telling them apart, at least in non-adverse conditions (Gandour, 1983). But this is not surprising given that the function of lexical tone is to *distinguish* words/morphemes that are otherwise identical to each other. Thus on the one hand only distinct tonal contrasts would have emerged and survived in a language, on the other hand, serviceable contrasts do not require precisions as high as in music. Such a natural design makes it possible for individuals who cannot sing in tune (10-15% of the population: Dalla Bella et al., 2007; Pfordresher et al. 2010) or detect out-of-tune singing (about 4% of the population: Kalmus & Fry, 1980; Nan et al., 2010) to still function normally in speech communication. The nature of speech therefore determines that speech is not about forms that are analogous to musical notes, but about the functions behind the forms, and so it is function-driven (Patel, 2008).

Lessons from phonemes

Functions in speech, however, are often elusive. Is it really possible to make a rigorous pursuit of functions? The answer is definitely yes, because functional approaches are not strangers to us. One of the most significant progress in linguistics, i.e., the development of the notion of phoneme, has actually established a basic principle of functional approaches. Based on this principle, a phonemic contrast is established, first and foremost, based on its ability to distinguish words or grammatical functions. Phonetic variants that do not distinguish one word from another, e.g., the many variety of /r/ in English, are considered as allophones rather than different phonemes. Thus lexical contrast is the *defining property* rather than *subordinate* or *accompanying property* of phonemes.

Another important aspect of the phonemic notion, which is not frequently highlighted, is that phonemes are not meaning carriers themselves, but serve only to distinguish lexical items from one another. The specific lexical meanings are morphologically rather than phonologically defined, and only occasionally, e.g. in the case of onomatopoeia, is there any direct link between phonetic form and lexical meaning. This is very different from the notion of "intonational meaning", according to which components of intonation are directly meaningful (cf. Ladd, 2008 for a review).

Following the phonemic tradition, then, the differences in form that matter phonologically would be those that make functional contrasts, rather than any observable differences, even if the observation is experimental. But if this is the case, there needs to be serious rethinking about many prosodic patterns that have been deemed "phonological" or important. These may include, e.g., *pitch accent*, *phrase accent* and *boundary tone* in the AM theory of intonation (Gussenhoven, 2004; Ladd, 2008; Pierrehumbert, 1980), patterns that are considered to be important for the description and teaching of intonation, e.g., *nucleus*, *head*, *pre-head* and *tail* in the British nuclear tradition (Halliday, 1967; Kingdon, 1958; O'Connor & Arnold, 1961; Palmer, 1922), rhythmic patterns such as *stress-timing*, *syllable-timing* and *mora-timing* (Ramu et al., 1999; Port, 2003; Warner & Arai, 2001), *prosodic hierarchy* in terms of *prominence* (Beckman, 1996; Selkirk, 1984), etc. In each of these cases, the critical question one may want to ask is, are the proposed categories primarily defined by the functions they serve, or is the functional aspect of the categories treated only as secondary?

Functional ≠ categorical

An important bias that is likely introduced by the phonemic tradition is the primacy given to categorical contrasts. The issue is not only about the laboratory phenomenon of categorical perception, which has already been met with questions (Fujisaki & Kawashima, 1971; Schouten et al., 2003), but also the notion that all linguistic or communicational contrasts have to be

categorical (Ladd, 2008; Liberman & Pierrehumbert, 1984). What needs to be recognized is that probably due to fluid vocabulary changes as well as phonological crowding (as exemplified by the rapid abstraction and categorization of newly invented sign languages over just a few generations: Pietrandrea, 2002), it is difficult to maintain iconicity in lexical construction, especially that involving gradience. For instance, it is hard to imagine representation of object size with a gradient change in vowel colour. For non-lexical functions, however, it is possible to represent gradient contrasts. For example, there is much evidence, though further research is still needed, that the demarcation/grouping function, as marked by F_0 height and domain-final lengthening, is gradient rather than categorical (Wagner, 2005; Byrd & Krivokapić, 2006). Therefore, categoricalness per se probably should not be used as the benchmark for determining whether a contrast is functional.

The need to establish mechanisms

Equally important as the need to focus on functions is the necessity to pursue the underlying mechanisms that encode the functions. This is because without establishing plausible mechanisms, the link between observed forms and underlying functions remains incomplete. Establishing a continuous link is not easy, however. In the case of prosody, for example, at least three degrees of separation between surface prosodic forms and the communicative functions they encode can be identified, namely, *articulatory constraints*, *target reassignment* and *parallel encoding* (Xu, 2004a). Each degree of separation actually involves a different set of mechanism. At the articulatory level are mechanisms like *target approximation*, *target-syllable synchronization*, *cross-boundary state transfer*, *anticipatory dissimilation*, *post-L bouncing*, *vowel intrinsic F_0*, *consonantal perturbation*, etc. (see Xu, 2006 for a review). Because of these mechanisms, directly observed surface acoustic forms can *never* fully resemble the underlying phonetic targets used to encode the functions.

At the level of target assignment, where an underlying pattern is determined for each syllable, there are often language-specific rules that change the targets depending on factors like phonetic context and communicative functions. The most striking example is the tone sandhi phenomenon in many tone languages (Chen, 2000). In Mandarin, for instance, the Low tone has a rising tail when produced in isolation and sometimes sentence-finally. But when it is followed by any other tone the rising tail is missing, and its absence has no plausible articulatory explanations (Xu, 2004b). Also, the Low tone changes into a Rising tone when followed by another Low tone, and again there are no plausible articulatory explanations. More recently, it is shown that target reassignment happens also in English intonation, where the underlying pitch target

associated with a stressed syllable varies across high, fall and rise depending on its position in word and sentence, whether it is in focus, and whether the sentence is a statement or question (Liu & Xu, 2007; Xu & Xu, 2005). Target reassignment, whenever it occurs, makes it difficult to directly recognize functionally relevant targets. But the recognition is not impossible if we employ methods that are sufficiently sensitive.

Beyond articulatory constraints and target assignment, there are also mechanisms that make it possible for multiple functions to be simultaneously encoded. Such parallel encoding of multiple functions, however, further obscures the link between any specific functions and the directly observable prosodic forms. To reveal the actual link for each function, it is necessary to conduct controlled experiments (e.g., Cooper et al., 1985; Eady & Cooper, 1986; Eady et al., 1986; Liu & Xu, 2007; Pell, 2001; Wagner, 2005; Xu, 1999; Xu & Xu, 2005). The observation of such multi-dimensional coding is what has inspired the parallel encoding and target approximation model (PENTA) (Xu, 2005).

Needless to say, any proposed mechanisms, including the ones just mentioned, are open for debate, as scientific hypotheses should be. What is important is to recognize the need to actively pursue them rather than being easily satisfied by seeming descriptive adequacy. Also, just as importantly, once recognized, mechanisms should always be taken into consideration whenever they may apply rather than being used only when convenient.

Potential benefit of function- and mechanism oriented approaches

Functions and mechanisms are not only of theoretical interest or beneficial only within a particular area of research, but they may also have broader impacts. In the case of prosody, for example, better understanding of functions and mechanisms may make the research findings more relevant for language teaching, speech technology, the relation between prosody and other levels of speech and the relation between linguistics and other disciplines. In regard to language teaching, the nuclear tone tradition was actually first developed as a tool for teaching English intonation (Palmer, 1922). However, after being promoted by generations of authors (e.g., Brazil, 1980; Cruttenden, 1997; Crystal, 1969; Halliday, 1967; Kingdon, 1958; O'Connor & Arnold, 1961; Palmer, 1922), its effectiveness in teaching English intonation is still yet to be demonstrated (Atoye, 2005). In this tradition, priority is given to detailed description of melodic contours of intonation in terms of nucleus, head, pre-head and tail. Assuming that the descriptions are reasonably accurate, why aren't they helpful for the students? Given that the communicative functions of these components are not clearly defined (other than that the nucleus is partially equivalent to focus), it would be difficult for students to learn when to use which pattern.

More importantly, not functionally-defined also means that each described melodic pattern may carry multiple functions, due to parallel encoding as discussed above. Teaching such confounded patterns is unlikely to facilitate the learning of true regularities in intonation.

In speech technology, it has been long desirable to significantly improve the prosody of synthesis, with the goal to generate highly natural as well as expressive speech. But the success has been limited so far. From a functional and mechanistic point of view, this is not surprising, because naturalness and expressiveness are not abstract properties. Truly natural and expressive speech, by definition, has to convey rich and appropriate communicative meanings, and should do so in a manner that resembles the human articulation process. Thus improved understanding of functions and mechanisms may therefore help to move speech technology forward toward being able to produce synthetic speech that is truly expressive and natural. Some efforts in this direction have already been made (Bailly & Holm, 2005; Prom-on et al., 2009).

The mechanistic-functional view of speech may also improve our understanding of the link between different levels of linguistic processing. For example, a long-standing issue in prosody is how prosodic structures are linked to syntactic structures. Some accounts favour close links (Chomsky & Halle, 1968; Selkirk, 1984), while others favour relative independence (Beckman, 1996; Shattuck-Hufnagel & Turk, 1996). From a functional perspective, however, the issue can be viewed in an entirely different light. That is, because speech is about exchanging information, both syntax and prosody should be for that purpose. But for the sake of information transmission, if a function is already syntactically coded, there is no need to also encode it prosodically, and vice versa, unless of course, *redundancy* of coding is favoured (Assmann & Summerfield, 2004). Redundant coding is indeed found for various functions. For example, focus is often marked by both syntactic and prosodic means (Féry et al., 2010). Likewise, boundary or grouping information is also likely encoded by both syntactic/semantic and prosodic means (Wagner, 2005). On the other hand, the specifics of such redundancy do not have to be universal. For example, recent research shows that post-focus compression is a highly effective prosodic cue of focus (Chen et al., 2009), but many languages do not have post-focus compression (Xu, 2011; Zerbian et al., 2010). This new finding has now led to hypotheses that may link cross-linguistic distributions of post-focus compression to recent findings in population genetics (Xu, 2011).

Conclusion

This paper has presented arguments for adopting a new perspective in linguistic research with gives high priority to the pursuit of communicative functions and the underlying mechanisms. It is argued that many previously reported phenomena may be viewed in a very different light in this perspective, which may actually enable us to quickly hone in on issues that really matter in speech. In addition to theoretical advantages of there are also potential benefits of adopting function-mechanism-oriented approaches.

References

Assmann, P.F., Summerfield, A.Q. 2004. The perception of speech under adverse conditions, in: S. Greenberg, W.A. Ainsworth, A.N. Popper, R. Fay (Eds.), Speech Processing in the Auditory System. Springer-Verlag, New York.

Atoye, R. O. 2005. Non-native perception and interpretation of English intonation. Nordic Journal of African Studies 14, 26-42.

Bailly, G., Holm, B. 2005. SFC: a trainable prosodic model. Speech Communication 46, 348-364.

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

Bolinger, D. 1986. Intonation and its parts: melody in spoken English. Stanford University Press, Palo Alto.

Brazil, D.M. Coulthard, M. and Johns, C., 1980. Discourse Intonation and Langguage Teaching. Longman, London.

Byrd, D., Krivokapić, J., Lee, S. 2006. How far, how long: On the temporal scope of phrase boundary effects. Journal of the Acoustical Society of America 120, 1589-1599.

Chen, M.Y. 2000. Tone Sandhi: Patterns across Chinese Dialects. Cambridge University Press, Cambridge, UK.

Chen, S.-w., Wang, B., Xu, Y. 2009. Closely related languages, different ways of realizing focus. Proceedings of Interspeech 2009, Brighton, UK**:** 1007-1010.

Chomsky, N., Halle, M. 1968. The Sound Pattern of English. Harper & Row, New York.

Cooper, W.E., Eady, S. J., Mueller, P.R. 1985. Acoustical aspects of contrastive stress in question-answer contexts. Journal of the Acoustical Society of America 77, 2142-2156.

Cruttenden, A. 1997. Intonation. Cambridge University Press, Cambridge.

Crystal, D. 1969. Prosodic Systems and Intonation in English. Cambridge University Press, London.

Dalla Bella, S., Giguère, J.-F., Peretz, I. 2007. Singing proficiency in the general population. Journal of the Acoustical Society of America 121, 1182-1189.

Eady, S. J., Cooper, W. E. 1986. Speech intonation and focus location in matched statements and questions. Journal of the Acoustical Society of America 80, 402-416.

Eady, S.J., Cooper, W.E., Klouda, G.V., Mueller, P.R., Lotts, D.W. 1986. Acoustic characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. Language and Speech 29, 233-251.

Féry, C., Skopeteas, S., Hornig, R. 2010. Cross-linguistic comparison of prosody, syntax and information structure in a production experiment on localising expressions. Transactions of the Philological Society 108(3), 329-351.

Fujisaki, H., Kawashima, T. 1971. A model of the mechanisms for speech perception: Quantitative analysis of categorical effects in discrimination. Annual Report of the Engineering Research Institute (Faculty of Engineering, University of Tokyo) 30, 59-68.

Gandour, J. 1983. Tone perception in Far Eastern languages. Journal of Phonetics 11, 149-175.

Grabe, E. 1998. Pitch accent realization in English and German. Journal of Phonetics 26, 129-143.

Gussenhoven, C. 2004. The Phonology of Tone and Intonation. Cambridge University Press.

Halliday, M.A.K. 1967. Intonation and Grammar in British English. Mouton, The Hague.

Hermes, D.J., Rump, H.H. 1994. Perception of prominence in speech intonation induced by rising and falling pitch movements. Journal of the Acoustical Society of America 96, 83-92.

Hirst, D.J. 2005. Form and function in the representation of speech prosody. Speech Communication 46, 334-347.

Kalmus, H., Fry, D.B. 1980. On tune deafness (dysmelodia): Frequency, development, genetics and musical background. Annals of the Human Genetics 43, 369-382.

Kingdon, R. 1958. The groundwork of English intonation. Longman, London.

Kochanski, G., Grabe, E., Coleman, J., Rosner, B. 2005. Loudness predicts prominence: Fundamental frequency lends little. Journal of the Acoustical Society of America 118, 1038-1054.

Kohler, K. 2005. Timing and Communicative Functions of Pitch Contours. Phonetica 62, 88-105.

Ladd, D.R., Faulkner, D., Faulkner, H., Schepman, A. 1999. Constant "segmental anchoring" of F0 movements under changes in speech rate. Journal of the Acoustical Society of America 106, 1543-1554.

Liberman, M., Pierrehumbert, J. 1984. Intonational invariance under changes in pitch range and length, in: M. Aronoff and R. Oehrle (Eds.), Language Sound Structure. M.I.T. Press, Cambridge, Massachusetts**:** 157-233.

Liu, F., Xu, Y. 2007. Question intonation as affected by word stress and focus in English. Proceedings of The 16th International Congress of Phonetic Sciences, Saarbrücken**:** 1189-1192.

Loukina, A., Kochanski, G., Rosner, B., Keane, E., Shih, C. 2011. Rhythm measures and dimensions of durational variation in speech. Journal of the Acoustical Society of America 129, 3258-3270.

Nan, Y., Sun, Y., Peretz, I., 2010. Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. Brain 133, 2635-2642.

O'Connor, J.D., Arnold, G.F. 1961. Intonation of Colloquial English. Longmans, London.

Palmer, H.E. 1922. English Intonation, with Systematic Exercises. Heffer, Cambridge.

Patel, A.D. 2008. Music, Language, and the Brain. Oxford University Press, New York.

Pell, M.D. 2001. Influence of emotion and focus on prosody in matched statements and questions. Journal of the Acoustical Society of America 109, 1668-1680.

Pfordresher, P.Q., Brown, S., Meier, K., Belyk, M., Liotti, M., 2010. Imprecise singing is widespread. Journal of the Acoustical Society of America 128, 2182-2190.

Pierrehumbert, J., Hirschberg, J., 1990. The meaning of intonational contours in the interpretation of discourse, in: P. R. Cohen, J. Morgan and M. E. Pollack (Eds.), Intentions in Communication. MIT Press, Cambridge, Massachusetts**:** 271-311.

Pierrehumbert, J. 1980. The Phonology and Phonetics of English Intonation. Ph.D. dissertation, MIT, Cambridge, MA. [Published in 1987 by Indiana University Linguistics Club, Bloomington].

Pietrandrea, P. 2002. Iconicity and Arbitrariness in Italian Sign Language. Sign Language Studies 2, 296-321.

Port, R.F. 2003. Meter and speech. Journal of Phonetics 31, 599–611.

Prom-on, S., Xu, Y., Thipakorn, B., 2009. Modeling tone and intonation in Mandarin and English as a process of target approximation. Journal of the Acoustical Society of America 125, 405-424.

Ramus, F., Nesporb, M., Mehlera, J. 1999. Correlates of linguistic rhythm in the speech signal. Cognition 73, 265-292.

Rump, H. H., Collier, R. 1996. Focus conditions and the prominence of pitch-accented syllables. Language and Speech 39, 1-17.

Schafer, A. J., Speer, S. R., Warren, P., White, D. 2000. Intonational disambiguation in sentence production and comprehension. Journal of Psycholinguistic Research 29, 169-182.

Schouten, B., Gerrits, E., Hessen, A. v. 2003. The end of categorical perception as we know it. Speech Communication 41, 71-80.

Selkirk, E. 1984. Phonology and syntax: the relation between sound and structure. MIT Press, Cambridge, Mass.

Shattuck-Hufnagel, S., Turk, A.E. 1996. A Prosody Tutorial for Investigators of Auditory Sentence Processing. Journal of Psycholinguistic Research 25(2), 193-247.

Shen, X.S. 1993. The use of prosody in disambiguation in Mandarin. Phonetica 50, 261-271.

Terken, J. 1991. Fundamental frequency and perceived prominence of accented syllables. Journal of the Acoustical Society of America 89, 1768-1776.

Wagner, M. 2005. Prosody and Recursion. Ph.D. Dissertation, Massachusetts Institute of Technology.

Warner, N., Arai, T. 2001. Japanese Mora-Timing: A Review. Phonetica 58, 1-25.

Xu, Y., Xu, C. X. 2005. Phonetic realization of focus in English declarative intonation. Journal of Phonetics 33, 159-197.

Xu, Y. 1999. Effects of tone and focus on the formation and alignment of F0 contours. Journal of Phonetics 27, 55-105.

Xu, Y. 2004a. Separation of functional components of tone and intonation from observed F0 patterns, in: G. Fant, H. Fujisaki, J. Cao, Y. Xu (Eds.), From Traditional Phonology to Modern Speech Processing: Festschrift for Professor Wu

Zongji's 95th Birthday. Foreign Language Teaching and Research Press, Beijing: 483-505.

Xu, Y. 2004b. Understanding tone from the perspective of production and perception. Language and Linguistics 5, 757-797.

Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. Speech Communication 46, 220-251.

Xu, Y. 2006. Principles of tone research. Proceedings of Second International Symposium on Tonal Aspects of Languages, La Rochelle, France: 3-13.

Xu, Y. 2011. Post-focus compression: Cross-linguistic distribution and historical origin. Proceedings of The 17th International Congress of Phonetic Sciences, Hong Kong.

Zerbian, S., Genzel, S., Kügler, F., 2010. Experimental work on prosodically-marked information structure in selected African languages (Afroasiatic and Niger-Congo). Proceedings of Speech Prosody 2010, Chicago: 100976:1-4.