

# In Defense of Lab Speech in Prosody Research

Yi Xu

University College London, UK

yi.xu@ucl.ac.uk

## Abstract

Lab speech has often been described as unnatural, overly clear, over planned, monotonous, lack of rich prosody, and lack of emotions. Along with this view is a growing popularity for directly examining spontaneous speech for the sake of understanding spontaneous speech, especially in regard to its prosody. In this paper I argue that few of the stereotyped characteristics associated with lab speech are warranted. Instead, the quality of lab speech is a design issue rather than a matter of fundamental limitation. More importantly, because it is controlled, the potential contribution of lab speech to our understanding of the nature of human speech far outweighs that of spontaneous speech.

## 1. Introduction

As speech research advances, there is a growing interest in aspects of speech beyond lexical phonemes like consonants, vowels and tones. In pursuit of this interest, many turned to spontaneous speech to look for answers. A widespread view is that only by directly examining spontaneous speech can we understand the nature of everyday speech. Accompanying this view is the increasingly popular idea that the so-called “lab speech” is grossly inadequate for shedding light on the richness of spontaneous speech. In this paper, I will argue that much of this belief is based on neglect of the literature, misconceptions about scientific inquiry, lack of imagination or simply failure to think things through.

## 2. Terminology

By a broad definition, lab speech refers to speech that is recorded in the laboratory, usually in the form of reading aloud scripts that are pre-composed. However, the term lab speech is often used to refer to a stereotyped speech such as:

Say *hid* again.  
Say *heed* again.  
Say *hood* again.

where the italicized words are the ones under scrutiny. But in fact, this type of lab speech is already a big improvement over earlier recordings in which syllables are recorded in isolation [31].

The progress from isolated vowel/words to vowels in a controlled syllable frame in a carrier sentence actually highlights the possibility of improvements in designing lab speech materials. But such possibility is typically ignored when people mention lab speech as a bad name. A more precise definition of lab speech should be something like *speech recorded under experimental control*, which more accurately represents the nature of lab speech.

Spontaneous speech, according to Beckman [4:7], is “speech that is not read to script”. She further distinguishes between ten different types of spontaneous speech recordings,

ranging from *unstructured narrative* to *instruction monologues*. The dividing line between lab speech and spontaneous speech can sometimes be blurred. For example, even when recording unscripted speech, certain levels of controls can be implemented. In what is referred to as *instruction monologues*, the speaker is asked to instruct a real or imaginary silent listener to perform a task. With this technique, some control over both content words and syntactic structure can be achieved [4]. To the extent the level of control is achieved, this type of speech could be labeled as lab speech as well. But, just as one would expect, the control is said to be achieved at the expense of naturalness [4].

## 3. Myths about lab speech

There are many myths about lab speech that are being passed around in the speech science community. But few of them are explicitly stated in peer-reviewed publications. They nevertheless have impacts on the way we conduct speech research. Although many researchers still use lab speech in their studies, they often do so apologetically, and are constantly thinking of ways to incorporate spontaneous speech into their research. In the following I will list a few what I believe are the most popular ideas about lab speech, and explain why they are actually just myths. Not all of these ideas are taken seriously by everyone, however, because some of them are so obviously false. But the more “credible” ones are in fact often closely related to the more simplistic ones, and it is thus important to point out their intimate relations.

### 3.1. Lab speech is slow and careful

This is probably one of the least sustainable myths. But many other myths are closely related to it. Speaking rate, as a matter of fact, is one aspect of speech that is among the most easily controlled in the laboratory. Numerous studies have been conducted in which speaking rate is systematically controlled, ranging from those that specifically look at the limits on the speed of articulation [21, 38, 53] to those that examine the effect of speaking rate on various phonetic aspects of speech [1, 7, 14, 17, 18, 20, 22, 23, 28, 33, 34, 36, 48, 50, 54].

Also there are different methods of manipulating speaking rate in the lab. The most straightforward is to simply ask speakers to speed up or slow down. While it is not easy to aim at a particular speaking rate as measured by, say, number of syllables per second, it is very easy to have untrained subjects speak at 2-3 different rates. My personal experience is that it is only difficult sometimes to make people speak very slowly without losing control over the aspect of speech under scrutiny. For example, in [54] we had to use only two speaking rates: normal and fast, lest the speakers would often insert pauses when producing focus at a very slow rate. The second strategy is to instruct subjects to speak casually or formally, or clearly or intimately, so as to elicit different speaking rates [29, 30]. Yet another way of controlling the rate of specific phonetic units is to control for local rate. For

example, in Mandarin, the middle syllable of a trisyllabic word is often spoken at a much faster rate than the surrounding syllables. Such local variability in rate has been explored in [37, 46, 50].

Also, controlling speaking style in the lab could allow us to separate variations due to speaking style and those due to other factors, such as speaking rate, as mentioned earlier. This has been done in [22].

Most importantly, the laboratory manipulation of speaking rate is so effective that some of the phenomena allegedly only occurring in spontaneous speech have been elicited in the lab. They will be discussed next, as they are also relevant for the myth about clarity of lab speech.

### 3.2. Lab speech is clear and articulate

This is closely related to the slowness myth, but somehow a little more sensible. That is, regardless of whether speaking rate is controlled for, speech recorded in the lab may tend to be clear and articulate. This is probably because, being in a lab, and asked to speak from a script, it is natural for speakers to speak clearly, just as they would when speaking to a foreigner or in front of a microphone for a formal occasion, or just reading aloud text in a classroom. But this kind stylistic tendency can be controlled. Speakers can be instructed to speak either more or less formally [13, 30, 32] and they do not seem to have much difficulty following such instructions. In fact, for some experimental purposes, we have to instruct subjects not to slur while trying to speak naturally [52]. It is a pure myth that everyone would uncontrollably speak in a careful manner as soon as they are in front of a microphone. There might be some people like that. But I have yet to encounter one in my own research.

Speakers' flexibility in controlling their own speaking style has made it possible for researchers to manipulate their speech along the dimension of clear versus casual in quite a few studies [13, 29, 30, 32, 41]. In an ongoing study, we have successfully elicited samples of syllable contractions from nonsense words embedded in meaningful sentence frames in Taiwan Mandarin, i.e., the merger of two or more syllables into one, which is generally believed to be characteristic of only casual speech [8].

Most importantly, controlling speaking style in the lab would allow us to separate variation due to speaking style and due to other factors, such as speaking rate, as mentioned earlier [22].

### 3.3. Lab speech is unnatural

This may be one of the most readily conceived characteristics of lab speech, because it seems to contain an element of truth, i.e., scripted speech, by definition, is non-spontaneous. And non-spontaneous is the opposite of being natural. But it is first of all important to point out something obvious. That is, even the most stereotypical lab speech is still human speech produced by real speakers. Regardless of what the speakers are asked to do in an experiment, their performance is based on their naturally acquired ability to speak, and is therefore a reflection of what they do everyday. It is not the case, for example, that they learn from the experimenter how to produce a vowel or a consonant or a tone, or they learn from the experimenter how to make focus or ask questions. They already know how to do those things, and that's why we want them in the laboratory in the first place. An experimental set-up only provides them with a situation in which the production

of certain aspects of speech is obligatory. Note that similar situations occur in natural conversations, too. The difference is that those situations are out of the control of the researchers.

Of course unnaturalness usually has a much broader connotation than not being spoken off the cuff. It is often associated with qualities such as careful, articulate, formal, etc. As discussed earlier, much of these characteristics are actually controllable in the lab and are thus not obligatory aspects of lab speech. On the other hand, it has to be admitted that lab speech can be unnatural in some sense. The more important question is, then, is the goal of the study compromised when the recorded speech sounds "unnatural"?

To answer this question, we can revisit some of the previous studies. In Peterson and Barney [31], vowels of American English are studied by embedding them in the syllable frame of h\_d and asking subjects to read a randomized list of 10 such words. This is perhaps one of the most stereotypical examples of lab speech, and so one may legitimately ask, what did the study miss by using lab speech? The purpose of the study is to understand the relationship between the perceptual identification of vowels and their acoustic properties. For this purpose the type of lab speech used, unnatural as it must have sounded, seems quite adequate. In [47], Mandarin tones are studied by asking native speakers to say bi-tonal combinations carried by the disyllabic sequence /mama/, which in turn are embedded in tonally balanced sentence carriers. Because the /mama/ sequences are mostly nonwords, the speech samples are by a narrow definition quite "unnatural". The goal of the study is to understand contextual tonal variations. Judging from the fact that the findings of the study have been corroborated multiple times in other studies of Mandarin [48, 49, 50] and other languages [15, 19, 25, 45] using either nonsense or meaningful materials, the goal of the study was not compromised. In general, whether anything is missing due to lack of naturalness is dependent on the match between the purpose of the study and the design of the experiment.

But the naturalness issue is still not fully resolved. If someone's speech in the lab does sound unnatural, we may further ask, what exactly has this person done to make his/her speech unnatural? Has he/she suddenly turned into a text-to-speech (TTS) system and started to generate speech like a TTS system does? Of course not. Or, is the person doing something that he/she has never done before in life? Probably not either. It is more likely that the person has spontaneously assumed a speaking style that seems to be appropriate for the occasion, i.e., reading aloud text in a place where serious business is going on. Such a style shift is not something artificial. Rather, it is a *natural* adjustment to the situation. From a functional point of view [51], such a style shift happens along an independent functional dimension which ranges from extremely casual to extremely formal. The nature of such a functional dimension no doubt requires research, as has been done indirectly in some recent investigations [29, 30, 32]. But the point here is that, as I have argued before, communicative functions are independent though parallel to each other [51]. The presence or absence of a particular function does not suppress the operations of other functions. Therefore, the so-called unnaturalness in lab speech is likely a manifestation of formality, which, though worth studying in its own right, would not invalidate the findings about other communicative functions based on lab speech, as has been shown in [6].

### 3.4. Lab speech is over planned

When speakers are asked to read aloud scripted texts during a recording session, naturally there is a possibility that they can plan for the whole utterance before starting to speak. But lab speech is not always fully planned. Just as we can manipulate the amount of information given to the listener in a perception experiment, we can also manipulate the amount and timing of information given to speakers to control their planning during production [39, 43]. Whalen [43], for example, controlled the amount of text subjects could see before starting to speak. By so doing he could examine the amount of anticipatory and carryover coarticulation that is plannable by the subjects. In Xu et al. [55], although subjects were given scripts of the sentences to be read aloud, their task was to imitate the exact manner with which the sentences were spoken by the model speaker. But because various parts of the speech of the model speaker was replaced with pink noise, subjects could not do much planning ahead of time. There can also be many other ways to control the amount of planning by the speaker. Whether and how such control is exerted, again, is a matter of experimental design which is closely related to the purpose of the research.

### 3.5. Lab speech is monotonous with impoverished prosody

This myth is apparently based on a poor understanding of what we already know about speech prosody. First, if we adopt a broad definition of prosody so that it covers any aspect of speech that is suprasegmental, lexical tones in languages like Mandarin would be included as part of prosody. Of course nobody in their right mind would deny that tones can be produced in the laboratory. Similarly, lexical stress in languages like English, which is also suprasegmental, is also easily observable in the lab. Secondly, even if we narrow down the definition of prosody to exclude anything lexical, there are still many prosodic patterns that are readily observable in lab speech. These include patterns associated with focus [5, 9, 49, 54], topic [24], grouping [40, 41], and sentence modality [11, 26, 27], etc.

What is important here is that prosodic patterns are used to encode various communicative functions [51]. When an experimental design does not include the right condition to make the encoding of a particular function obligatory, the associated prosodic pattern is not guaranteed to occur. Thus the lack of various particular prosodic patterns in many laboratory experiments is often either due to deliberate exclusion of those functions, or lack of proper methods to elicit them. But either way the issue is over how and how well an experiment is designed, not whether lab speech allows us to study prosody at all. Judging from the fruitful returns of so many studies, it certainly does.

### 3.6. Lab speech is emotionless

This is certainly untrue given that many studies have used speech samples with emotions enacted in the laboratory. Questions can no doubt be raised about the authenticity of the enacted emotions. But as will be argued later, to use anything that occurs naturally as an object of study, the first obstacle to overcome is the correct classification of that object. This makes emotions in spontaneous speech just as elusive as those in enacted speech. Again, however, because lab speech is controllable, the methods of eliciting emotions can be

continually improved, limited perhaps only by ethical concerns in some situations, e.g., those linked to extreme emotions. But again, similar restrictions may be applicable to spontaneous speech with extreme emotions as well.

### 3.7. Interim summary

The above discussion has shown that many myths about lab speech are not well-founded. In general, the characteristics attached to lab speech are related to the purpose of the study rather than to lab speech as a whole. When we want to understand vowels, consonants and tones, we have to be able to control the variation of these aspects of speech while keeping other aspects constant. What this means is that the non-manipulated aspects are left either in their neutral state, or in a state appropriate for the recording situation. But these other aspects can be also manipulated when the purpose of a study requires it. In particular, various prosodic functions can be specifically controlled, as has been done in many studies.

## 4. Spontaneous vs. lab speech

The appeal of spontaneous speech is that it can potentially make up for what is missing from lab speech [4]. Although we have seen that some popular ideas about what is missing from lab speech are actually based on myths, is it still possible that spontaneous speech can at least offer some more? My answer to this question starts from reconsidering the fundamental motivations of examining spontaneous speech.

One of the assumptions behind the drive to look at spontaneous speech is that science progresses by accumulating observations. Thus it is hoped that by looking at more and more samples of spontaneous speech, our knowledge about speech will keep improving. But as pointed out by Popper [35], no observation can be theory free or non-selective. All observations are selective and theory-laden. But theory-laden does not mean that observations are always driven by theories that are widely accepted or hotly contested. They could be based on theories that are formed “on the run”. For example, suppose we have no knowledge about the intonation of a particular language and we start by directly observing the  $F_0$  contours of the language. We may notice that there are prominent peaks and valleys in the  $F_0$  tracks. If we report our observations by summarizing the locations and sizes of those peaks and valleys, we may think that our report is free of any grand theories. That may be true. But such description of the intonation of this language is actually driven by our own petty theories formed as we made the observations. That is, we have assumed that, a)  $F_0$  peaks and valleys are *important events* in intonation, b) they are *direct correlates* of certain important linguistic categories, and c) what is obvious to the eye, e.g., peaks and valleys, is also obvious to the ear. Note that, each of these is actually a theoretical postulation, and as such they all need further assessment as to their validity.

How, then, do we assess a theory of speech, grand or petite as it may be? Do we dive into a spontaneous speech database and look for proof? If we do, how do we control the factors that may have contributed to the measurements we have taken? For example, we now know that the  $F_0$  contour of a syllable is determined not only by its tone or stress, but also by factors such as tonal context, focus, sentence type, topic, etc. [51]. But how do we control them in a spontaneous speech database? Anyone who has attempted to do so would attest that the task is extremely difficult, if not impossible. In

fact, to be able to find utterances that would fit the requirements of all the experimental conditions, the database would need to be almost infinitely large. In contrast, all these factors can be easily controlled under experimental conditions, as mentioned earlier.

Another motivation for looking at spontaneous speech is the belief that it is much richer than lab speech in terms of the variety of prosodic patterns. This may be true if by spontaneous speech we mean all the speech utterances produced by all speakers in a language community. By definition, a corpus of all the spontaneous utterances should indeed contain all the prosodic patterns. The problem is, no one can ever have access to such a corpus. Instead, real-life spontaneous speech corpora are all very limited in terms of the number of utterances as well as the types of prosodic patterns contained. More importantly, even if a particular prosodic event, say, focus, does occur in an utterance, to understand it, we need to compare it with another utterance in which focus is absent. But chances are that those utterances in the corpus which lack the equivalent focus are also different in terms of other factors, such as syllable structure, word structure, tonal context, sentence type, location in sentence, location in the paragraph, and so on. In fact, finding a single minimal pair in a spontaneous corpus that satisfies all the conditions is anything but trivial. Finding multiple pairs, as typically required by a controlled experiment, is close to impossible. And, to make things worse, even if a minimal pair happens to be found based on a particular set of conditions, chances are that it would be no longer valid as soon as a new condition is added. Given such limitations, it is very difficult to conduct a rigorous study using a spontaneous speech corpus.

A further difficulty with spontaneous speech is the problem of labeling. Any speech corpus needs to be carefully labeled for its internal elements before it can be subjected to research analysis. However, the labeling and analysis constitute an inherently circular process, as has been recognized [4:12]. That is, the labeling process assumes that we already know what and how to label, but the analysis process assumes that we still don't know the nature, the identity, or even the locations of those elements. This circularity problem is exacerbated if the labeling is done on the basis of direct observations. For example, in the ToBI convention of labeling intonation, pitch accents labels are attached to the visually prominent  $F_0$  peaks and valleys. Thus an analysis of the corpus based on these labels is virtually taking for granted the assumptions behind the labels, thus effectively treating a significant portion of the signal as not needing further analysis. Although this problem can be somewhat alleviated by doing what is suggested by Wightman [44], namely, to label only what you hear, we are still left with the assumption that the labelers *know what to listen for* in the uncontrolled speech utterances.

Of course, I would not go so far as saying that spontaneous speech corpora are useless. As collections of various natural patterns, they may be useful in motivating new hypotheses and raising questions about existing ones. But even on that ground I would also like to note that theoretical postulations do not actually need to be based on direct observations. This is because how a theory is initially conceived is irrelevant to science according to the Popperian view [35]. What is critical is that theories need to be tested through falsification. And it is the stringent requirements of

the falsification process that is hard for spontaneous speech corpora to meet.

At this point it might be helpful to take an excursion out of the field of speech communication to take a look at psychology for a debate that happened about 20 years ago over whether memory research should focus on "everyday memory" as opposed to laboratory memory. The debate occurred amidst a popular drive to study everyday memory in order to increase the ecological validity of memory research. That drive is not very unlike the current popular surge in speech research to study spontaneous speech in order to increase generalizability to everyday speech. But the problems with everyday memory are also not unlike those with spontaneous speech discussed above, as pointed out by Banaji and Crowder:

... the multiplicity of uncontrolled factors in naturalistic contexts actually prohibits generalizability to other situations with different parameters. The implication that tests in the real world permit greater generalizability is false once the immense variability from one real-world situation to another is recognized. [2:1189]

Because of such problems, the research with everyday memory has not been fruitful:

"No theories that have unprecedented explanatory power have been produced; no new principles of memory have been discovered; and no methods of data collection have been developed that add sophistication or precision." [2:1185]

I am not in a position to jump to conclusions about the fruitfulness of research based on spontaneous speech, as I have not yet done an exhaustive survey of the spontaneous speech literature. But I am quite sympathetic to the conclusions of Garner in a report to the US Office of Naval Research, as cited in [3:79]: "operational experimentation is more time consuming, far more expensive, and frequently cannot control experimental factors, so that as a practical matter it is very difficult to do operational experimentation which has a high degree of generality of prediction." Here operational experimentation means experimental manipulations in the operational field as opposed to in the laboratory. In the case of speech research, it may be about time for the field to do a similar assessment: Has the return been good with the vast amount of money spent on building and analyzing numerous spontaneous speech corpora?

## 5. Conclusions

Despite its increasing unpopularity, many of us are still looking at lab speech in our research on both the segmental and prosodic aspects of speech. But many of us are doing so with a guilty conscience, and frequently have to be apologetic about the speech materials that we have used. After examining the major complaints against lab speech, I have shown that virtually all of them are unfounded. It is not true that lab speech is uniformly slow and articulate, unnatural, over planned, monotonous with impoverished prosody, and emotionless. Rather, these characteristics are seen in some of the lab speech samples partly due to the purpose of the study, and partly due to the crudeness of experimental design in some cases, but never due to fundamental limitations of lab speech in general. I have argued in particular that naturalness itself may be related to degrees of formality, which is likely a

communicative function in its own right, and as such can be also studied in the laboratory.

I have also argued that although spontaneous speech corpora may allow us to make initial observations, true progress in our understanding of speech has to rely heavily on lab speech. This is because science progresses not by collecting more data, but by “hypothesis derivation from theory and hypothesis testing in the laboratory” [2:1192, 35]. Spontaneous speech can rarely allow us to fully control the factors that contribute to the phenomena we are interested in, which makes rigorous hypothesis testing difficult. The richness of spontaneous speech therefore may actually form impenetrable obstacles to true understanding. In contrast, experimental control in the lab allows us to make observations by manipulating the factors under investigation while keeping other factors constant. Observed variations can then be directly attributed to the manipulated factors. This is of course by no means an easy process, and the techniques we employ need constant update in order for us to gain increasingly better insights into the full complexity of speech in general, and prosody in particular. But marginalizing lab speech is clearly the wrong way to go.

## 6. References

- [1] Adams, S. G.; Weismer, G.; Kent, R. D., 1993. Speaking rate and speech movement velocity profiles. *Journal of Speech and Hearing Research* 36: 41-54.
- [2] Banaji, M. R.; Crowder, R. G., 1989. The bankruptcy of everyday memory. *American Psychologist* 44: 1185-1193.
- [3] Banaji, M. R.; Crowder, R. G., 1991. Some everyday thoughts on ecologically valid methods. *American Psychologist* 46: 78-79.
- [4] Beckman, M. E., 1997. A typology of spontaneous speech. In *Computing Prosody: Computational Models for Processing Spontaneous Speech*. Y. Sagisaka, N. Campbell and N. Higuchi. (eds.) New York: Springer Verlag: 7-26.
- [5] Botinis, A.; Fourakis, M.; Gawronska, B., 1999. Focus identification in English, Greek and Swedish. In *Proceedings of The 14th International Congress of Phonetic Sciences, San Francisco*. pp. 1557-1560.
- [6] Bruce, G.; Touati, P., 1992. On the analysis of prosody in spontaneous speech with exemplification from Swedish and French. *Speech Communication* 11: 453-458.
- [7] Caspers, J.; van Heuven, V. J., 1993. Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50: 161-171.
- [8] Cheng, C.; Xu, Y., 2008. Acoustic Features of Disyllabic Contraction in Mandarin. Presented at BAAP 08 Colloquium, Sheffield, UK. pp.
- [9] 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.
- [10] Crystal, T. H.; House, A. S., 1990. Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America* 88: 101-112.
- [11] 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.
- [12] Ferguson, S. H.; Kewley-Port, D., 2002. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America* 112: 259-271.
- [13] Gagne, J.-P.; Rochette, A.-J.; Charest, M., 2002. Auditory, visual and audiovisual clear speech. *Speech Communication* 37: 213-230.
- [14] Gandour, J., 1999. Effects of speaking rate on Thai tones. *Phonetica* 56: 123-134.
- [15] Gandour, J.; Potisuk, S.; Dechongkit, S., 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22: 477-492.
- [16] Garner, W. R., 1950. The validity of prediction from laboratory experiments to naval operational situations in the area of human engineering and systems research (Report No. 166-I-130). Baltimore, Johns Hopkins University, Institute for Cooperative Research.
- [17] Gay, T. J., 1968. Effect of speaking rate on diphthong formant movements. *Journal of the Acoustical Society of America* 44: 1570-1573.
- [18] Gay, T. J., 1978. Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America* 63: 223-230.
- [19] Gu, W.; Hirose, K.; Fujisaki, H., 2007. Analysis of Tones in Cantonese Speech Based on the Command-Response Model. *Phonetica* 64: 29-62.
- [20] Hirata, Y., 2004. Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 32: 565-589.
- [21] Janse, E., 2003. Word perception in fast speech: artificially time-compressed vs. naturally produced fast speech. *Speech Communication* 42: 155-173.
- [22] Krause, J. C.; Braida, L. D., 2004. Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America* 115: 362-378.
- [23] Kuo, Y.-C.; Xu, Y.; Yip, M., 2007. The phonetics and phonology of apparent cases of iterative tonal change in Standard Chinese. In *Tones and Tunes Vol 2: Experimental Studies in Word and Sentence Prosody*. C. Gussenhoven and T. Riad. (eds.) Berlin: Mouton de Gruyter: 211-237.
- [24] Lehiste, I., 1975. The phonetic structure of paragraphs. In *Structure and process in speech perception*. A. Cohen and S. E. G. Neebom. (eds.) Springer-Verlag: New York: 195-206.
- [25] Li, Y. J.; Lee, T., 2002. Acoustical F0 analysis of continuous Cantonese speech. In *Proceedings of International Symposium on Chinese Spoken Language Processing 2002, Taipei, Taiwan*. pp. 127-130.
- [26] Liu, F.; Xu, Y., 2005. Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* 62: 70-87.
- [27] Liu, F.; Xu, Y., 2007. Interaction of word stress, focus, and sentence type in English. *Journal of the Acoustical Society of America* 121, Pt. 2: 3199.
- [28] Miller, J. L.; O'Rourke, T. B.; Volaitis, L. E., 1997. Internal structure of phonetic categories: Effects of speaking rate. *Phonetica* 54: 121-137.
- [29] Moon, S.-J.; Lindblom, B., 1994. Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America* 96: 40-55.
- [30] Perkell, J. S.; Zandipour, M.; Matthies, M. L.; Lane, H., 2002. Economy of effort in different speaking conditions.

- I. A preliminary study of intersubject differences and modeling issues. *Journal of the Acoustical Society of America* 112: 1627-1641.
- [31] Peterson, G. E.; Barney, H. L., 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24: 175-184.
- [32] Picheny, M. A.; Durlach, N. I.; Braida, L. D., 1986. Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research* 29: 434-446.
- [33] Pind, J., 1995. Speaking rate, voice-onset time, and quantity: the search for higher-order invariants for two Icelandic speech cues. *Perception and psychophysics* 57: 291-304.
- [34] Pitermann, M., 2000. Effect of speaking rate and contrastive stress on formant dynamics and vowel perception. *Journal of the Acoustical Society of America* 107: 3425-3437.
- [35] Popper, K., 1959. *The Logic of Scientific Discovery* (translation of Logik der Forschung, 1934). London: Hutchinson.
- [36] Prieto, P.; Torreira, F., 2007. The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics* 35(4): 473-500.
- [37] Shih, C., 1993. Relative prominence of tonal targets. In Proceedings of The 5th North American Conference on Chinese Linguistics, Newark, Delaware, University of Delaware. pp. 36.
- [38] Tiffany, W. R., 1980. The effects of syllable structure on diadochokinetic and reading rates. *Journal of Speech and Hearing Research* 23: 894-908.
- [39] van Heuven, J. V., 2004. Planning in speech melody: production and perception of downstep in Dutch. In *On Speech and Language: Studies for Sieb G. Nootboom*. H. Quené and J. V. v. Heuven. (eds.) The Netherlands: LOT Occasional series by Utrecht University: 83-93.
- [40] Wagner, M. (2005). *Prosody and Recursion*. Ph.D. Dissertation. Massachusetts Institute of Technology.
- [41] Wang, M.; Xu, Y. 2008. F0 pattern and syllable grouping in Mandarin. This volume.
- [42] Weismer, G.; Berry, J., 2003. Effects of speaking rate on second formant trajectories of selected vocalic nuclei. *Journal of the Acoustical Society of America* 113: 3363-3378.
- [43] Whalen, D. H., 1990. Coarticulation is largely planned. *Journal of Phonetics* 18: 3-35.
- [44] Wightman, C. W., 2002. ToBI or not ToBI. In Proceedings of The 1st International Conference on Speech Prosody, Aix-en-Provence, France. pp. 25-29.
- [45] Wong, Y. W., 2006. Contextual Tonal Variations and Pitch Targets in Cantonese. In Proceedings of Speech Prosody 2006, Dresden, Germany. pp. PS3-13-199.
- [46] Xu, Y., 1994. Production and perception of coarticulated tones. *Journal of the Acoustical Society of America* 95: 2240-2253.
- [47] Xu, Y., 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61-83.
- [48] Xu, Y., 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55: 179-203.
- [49] Xu, Y., 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27: 55-105.
- [50] Xu, Y., 2001. Fundamental frequency peak delay in Mandarin. *Phonetica* 58: 26-52.
- [51] Xu, Y., 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46: 220-251.
- [52] Xu, Y., 2007. How often is maximum speed of articulation approached in speech? *Journal of the Acoustical Society of America* 121, Pt. 2: 3199-3140.
- [53] Xu, Y.; Sun, X., 2002. Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111: 1399-1413.
- [54] Xu, Y.; Xu, C. X., 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33: 159-197.
- [55] Xu, Y.; Xu, C. X.; Sun, X., 2004. On the Temporal Domain of Focus. In Proceedings of International Conference on Speech Prosody 2004, Nara, Japan. pp. 81-84.