Journal of Phonetics 38: 329-336 (Letter to the Editor) [Author version]

In Defense of Lab Speech

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 devoid of communicative functions, interactions and 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 allows systematic experimental control, lab speech is indispensable in our quest to understand the underlying mechanisms of human language. In contrast, although spontaneous speech is rich in various patterns, and so is useful for many purposes, the difficulty in recognizing and controlling the contributing factors makes it less likely than lab speech to lead to true insight about the nature of human speech.

1. Introduction

As speech research advances, there is a growing interest in aspects of speech beyond lexical contrasts due to consonants, vowels and lexical 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 letter, I argue that much of this belief is based on incomplete consideration of the literature, inadequate appreciation of scientific principles, lack of imagination or simply failure to think things through.

On a broad definition, lab speech refers to speech that is recorded in the laboratory, usually in the form of pre-composed scripts to be read aloud. 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 or even vowels are

recorded in isolation (e.g., Peterson & Barney, 1952), because the inclusion of a carrier allows the control of the effect of immediate phonetic context.

The progress from isolated vowels 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 lab speech is described as disadvantageous. A more precise definition of lab speech should be something like *speech sampled under experimental control*, which more accurately represents the nature of lab speech.

Spontaneous speech, according to Beckman (1997:7), 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 control 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 (Beckman, 1997). To the extent the level of control is achieved, this type of speech could be labeled as lab speech as well.

2. Myths about lab speech

There are many myths about lab speech in the speech science community. But few of them are explicitly stated in peer-reviewed publications (but see Rischel, 1992). 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 mention a few what I believe are the most popular characterizations of lab speech, and explain why they are actually unfounded stereotypes. Not all of these characterizations are taken seriously by everyone, because some of them are obviously untrue. But the more "credible" ones are in fact often closely related to the more simplistic ones, and it is thus important to point out the relations between them.

2.1. 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 (Janse, 2003; Adank & Janse, 2009; Tiffany, 1980; Xu & Sun, 2002) to those that examine the effect of speaking rate on various phonetic aspects of speech (Adams, Weismer & Kent, 1993; Caspers & van Heuven, 1993; Gandour, 1999; Gay, 1968, 1978; Hirata, 2004; Krause & Braida, 2004; Kuo, Xu & Yip, 2007; Miller, O'Rourke & Volaitis, 1997; Pind, 1995; Pitermann, 2000; Prieto & Torreira, 2007; Xu, 1998, 2001; Xu & Xu, 2005).

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 certain aspects of speech under scrutiny. For example, in Xu and Xu (2005) we had to

use only two speaking rates: normal and fast, because otherwise the speakers would often insert a pause when producing focus at a very slow rate. The second strategy is to instruct subjects to speak clearly or casually, or formally or intimately, so as to elicit different speaking rates (Moon & Lindblom, 1994; Perkell et al., 2002). 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 (Xu & Wang, 2009). Such local variability in rate has been explored in Shih (1993) and Xu (1994, 2001).

Most importantly, the laboratory manipulation of speaking rate is so effective that some of the phenomena allegedly occurring only 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.

2.2. Lab speech is clear and articulate

This is closely related to the slowness myth, although it is a little more realistic. 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 laboratory, 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 in 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 (Ferguson & Kewley-Port, 2002; Gagne, Rochette & Charest, 2002; Perkell et al., 2002; Picheny, Durlach & Braida, 1986; Xu & Wang, 2009) 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 (Xu, 2007). It is not true, in my own experience, that everyone would uncontrollably adopt a careful manner of speaking as soon as they are in front of a microphone.[i]

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 as mentioned above. In a recent 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 (Cheng & Xu, 2009).

Most importantly, controlling speaking style in the laboratory allows us to separate variations due to speaking style from those due to other factors, such as speech rate, as has been done in Krause & Braida (2004).

2.3. 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 seems to be the opposite of natural. But it is important to first determine what is meant by "natural". If naturalness is taken to mean reflecting human capabilities, all speech must be natural by definition. Indeed, even the most stereotyped lab speech is 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 an emphasis or ask questions. They already know how to do those things, and that's why they are invited into the laboratory in the first place.

But if "natural" is used to only refer to conversational speech or spontaneous monologues, many corpora, like the widely used TIMIT (Garofolo et al., 1993), Boston University Radio Speech Corpus (Ostendorf, Price & Shattuck-Hufnagel, 1996) and some others also available from the Linguistic Data Consortium (LDC, 2009), cannot be qualified as being natural, because they consist of only scripted monologues. As far as I know, however, these corpora are rarely characterized as unnatural, nor are they labeled as lab speech despite being read from scripts.

So, if a person's speech recorded in the lab does sound unnatural, it is neither because it is not conversational, nor because it is non-human. What, then, may have made the person's speech sound unnatural? Is it because s/he is doing something that has never been done in his/her life? Probably not. It is more likely that the person has spontaneously assumed a speaking style that is appropriate for the occasion, i.e., reading aloud text in a formal setting. Such a style shift is not artificial, but a rather *natural* adjustment to the situation. From a functional point of view (Barry, 1981; Hirst, 2005; Kohler, 2005; Xu, 2005), this style shift probably happens along a functional dimension that could be called *formality*. Assuming that communicative functions are independent of each other, the presence or absence of a particular function would not suppress the operations of other functions (Xu, 2005). Therefore, the stylistic unnaturalness in some lab speech would not invalidate the findings about other communicative functions based on lab speech. More importantly, the formality function itself can also be studied in the laboratory as has been done for other stylistic variations (e.g., Harnsberger, Wright & Pisoni, 2008; Moon & Lindblom, 1994; Perkell et al., 2002; Picheny et al., 1986).

2.4. 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 does not have to be 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 (van Heuven, 2004; Whalen, 1990). Whalen (1990), for example, controlled the amount of text subjects could see before starting to speak. By so doing the amount of anticipatory and carryover coarticulation that is plannable by the subjects could be examined. In Xu, Xu and Sun (2004), 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 were 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.

2.5. Lab speech is monotonous with impoverished prosody

This characterization is apparently based on an inadequate 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 few would deny that tones can be produced in the

laboratory. Similarly, lexical stress in languages like English, which is also suprasegmental, is easily observable in the lab as well. 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. Although it is debatable whether these patterns directly reflect various communicative functions (Hirst, 2005; Kohler, 2005; Xu, 2005, 2009) or a hierarchical prosodic structure (Beckman, 1996; Gussenhove, 2004; Ladd, 2008; Pierrehumber, 1980; Shattuck-Hufnagel & Turk, 1996), there is little doubt that they can be systematically elicited in the laboratory. Past studies have been able to systematically examine patterns associated with focus (Botinis, Fourakis & Gawronska, 1999; Cooper, Eady & Mueller, 1985; Xu, 1999; Xu & Xu, 2005), topic (Lehiste, 1975; Wang & Xu, 2006), grouping (Wagner, 2005; Xu & Wang, 2009), and sentence type (Eady & Cooper, 1986; Liu & Xu, 2005, 2007), to mention just a few.

An important lesson from these empirical studies is that 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. For example, an experiment examining only statements cannot reveal anything about how the interrogative function is encoded; and a study of interrogative function that does not systematically manipulate focus may not reveal how the two functions interact with each other (Eady & Cooper, 1986; Liu & Xu, 2005, 2007). 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 about 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, studying prosody with lab speech is certainly possible.

2.6. Lab speech is devoid of communicative functions

As an example of this widespread stereotype associated to lab speech, Rischel (1992:382, 383) stated that: “in spite of the programmatically empirical nature of phonetic research it is normal research strategy for all of us to use as our data not genuine, unmonitored speech, but on the contrary "lab speech" specimens that are explicitly devoid of any linguistic function for the speaker.”

The main reason for this kind of stereotype is that the “linguistic” or communicative function is not clearly defined. One of the most important functions of speech, for example, is to convey meanings through words, and word identity representation is therefore one of the basic linguistic functions. We know that such a function is achieved mainly by phonemic contrast, and even the most “unnatural” lab speech specimens are likely to have this function.

Of course what Rischel really meant by function is more likely to be pragmatic functions not symbolically represented in the text. But if we make those functions explicit we can see that many of them can be also easily elicited in the laboratory, as will be discussed in more detail in Section 5. In fact, as I will argue, unless a function is demonstrable in the laboratory, we cannot know for certain that it exists.

2.7. Lab speech is emotionless

This is untrue at least 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 (Scherer, 2003). 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 no less elusive than those in enacted speech. Again, however, because lab speech is controllable, the methods of eliciting emotions can be continually improved, limited perhaps only by ethnical concerns in some situations, e.g., those linked to extreme emotions.

2.8. Lab speech is devoid of interactive functions

The image of lab speech as involving only monologue reading of isolated words and sentences may have given rise to the perception that it is also severely inadequate in studying anything interactive in speech, such as turn taking, information structure, sociophonetic variations. But just as the other myths, this one is also based on insufficient consideration of what has been done in lab-based research in these areas. Regarding turn taking, Schafer et al. (2000) employed a cooperative game task, in which two speakers used utterances from a predetermined set of *scripted* sentences to negotiate moves around gameboards. For information structure, research has been conducted to study focus and topic in various languages (Cooper et al., 1985; Féry & Kügler, 2008; Lehiste, 1975; Wang & Xu, 2006; Xu, 1999; Xu & Xu, 2005). For sociophonetics, there have been experimental studies looking at factors that affect accent change (Evans & Iverson, 2004, 2007).

An issue related to interactive communications that has not been fully explored is the nature of the triggers in interactive functions. For example, a prosodic focus is triggered by a context that renders a particular part of the target sentence important. To answer the question "Who did she play in the movie?", the speaker would say "She played the PRINCESS in the movie". However, it is interesting to note that the prosodic focus is triggered whether the context question is a replay of a pre-recorded audio (Xu & Xu, 2005), spoken live by another interlocutor (Xu & Wang, 2009), or spoken by the speaker him/herself (Xu, 1999). It seems that the nature of interaction for focus is *informational* rather than *inter-personal*, which may differ from certain other functions for which the triggers are more social. Again, however, only well controlled experiments can lead to a better understanding of issues like this.

Finally, as mentioned earlier, even unscripted speech can be controlled to some extent in hybrid paradigms such as the map task and similar protocols (Edlund & Heldner, 2005; Nakajima & Allen, 1993). But even here it is my judgment that the level of control is commensurate with the level of certainty one can achieve in regard to the phonetic details corresponding to specific communicative functions.

2.9. Myths no more

The above discussion has shown that popular 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 in general. 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. Thus 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 so requires. In particular, various prosodic functions can be specifically manipulated, as has been done in many studies.

3. Spontaneous vs. lab speech

The above discussion probably is still not enough to resolve an essential question many of us may have in mind: Wouldn't it be best to study spontaneous speech *directly* in order to understand it? Before answering this question, it is necessary to first clarify two points. The

first is that spontaneous speech is not equivalent to corpus speech. This is because, as mentioned earlier, many speech corpora actually consist of only read speech (sentences, paragraphs or whole stories, radio or TV news, etc.), and thus the speech samples they contain are by no means spontaneous. How they differ from more prototypical lab speech is having a lower level of control implemented in the design of the texts. The second point is that the often-assumed dichotomy between naturally-occurring and laboratory speech as an object of study (e.g., Rischel, 1992:380; Silverman et al., 1992:867) is a false one, because lab speech is never the real object of study. No matter how unnatural the examined speech samples may be, the real objective is always to understand the kind of speech that occurs outside the lab (or general human capabilities and how they relate to performance in specific situations). This is no different from chemistry and biology where what happens in the test tube is never the real object of research, but a reenactment of what is happening in the wild, except with all the known factors under laboratory control. So the real dichotomy is whether it is better to observe what actually happens in natural environment or reenactments of the same phenomenon in the laboratory. I will discuss this question from several perspectives.

3.1. Generality

One of the main motivations for looking at spontaneous speech is the belief that it is much richer than lab speech in terms of the variety of prosodic patterns, and therefore findings based on spontaneous speech have greater generality than those based on lab speech. But if we take generality to mean the extent to which conclusions of a research study can be extended (Hedge, 1994), or more specifically the range of conditions under which a demonstrated cause-effect relationship holds good (Schlosser, 1994), spontaneous speech is problematic. First of all, any real-life spontaneous speech corpus, no matter how big it is, is always limited in terms of the number of utterances as well as the types of prosodic patterns it contains. To study a particular prosodic phenomenon, e.g., focus, in such a corpus, we need to compare *minimal pairs* of utterances in which focus is either present or absent. But chances are that those utterances are also different in terms of other factors, such as syllable structure, word structure, tonal context, sentence type, location in sentence, location in 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 for a controlled experiment, is next to impossible. And, to make things worse, even if a minimal pair happens to be found based on a particular set of conditions, it is likely to no longer qualify as a minimal pair as soon as a new condition is added. For example, what if we want to avoid the confound of speaker differences by looking only at minimal pairs spoken by the same speaker? What if we also want to control for intra-speaker variation by observing several occurrences of the same utterance by a single speaker? Both are difficult to achieve with a spontaneous speech corpus. Given such limitations, how can we know for certain if any cause-effect relation we want to extract from a spontaneous speech corpus is not due to factors not controlled for? Without such certainty, how can we be confident that our conclusions are generalizable to other situations?

The lack of systematic data in speech corpora is known in computational modeling as the Large Number of Rare Events (LNRE) problem, and various algorithms have been proposed to handle it (Möbius, 2003; van Santen, 1993). However, the problem has not yet been fully removed, and as pointed out by (Möbius, 2003:69), “the most promising avenue of research is to increase the coverage of speech databases by carefully defining the linguistic and phonetic criteria that the database should meet”. A possible strategy along this line that has seldom been adopted is to directly use lab speech as training corpora, as done in our recent modeling and synthesis of prosody in English and Mandarin (Prom-on, Xu & Thipakorn,

2009). The lesson we have learned is that not only are lab speech databases usable as training corpora, but also they may be superior to less controlled corpora (unscripted or scripted). This is because no matter how rich a corpus may be in terms of speech phenomena, the training process can only reproduce those communicative functions that have been systematically represented, while everything else can only contribute to the averages, and averaged speech will inevitably sound neutralized.

Of course, the generality question can be asked of lab speech as well, that is, how can we be sure that patterns found in lab speech are generalizable to spontaneous speech? One way to find out is to directly check if these patterns match those found in spontaneous speech. Although this has not been done frequently, at least one study specifically examined whether prosodic patterns reported for lab speech in Swedish and French could be also found in spontaneous speech (Bruce & Touati, 1992). Their conclusion was that for Swedish no fundamental differences exist between the two types of speech, and for French focal accent and contrast in pitch range could account for typical prosodic means used during political debate.

3.2. Data- vs. theory-driven

Another reason for favoring spontaneous speech is that some of us feel uncomfortable about theory-driven research. They feel reluctant to be biased by a particular theory and prefer to avoid using lab speech whose design is likely to be tinted by one theory or another. But as pointed out by Popper (1959), no observation can be theory free or non-selective. All observations are selective and theory-laden. Note that being theory-laden does not mean that the observations are necessarily driven by a theory that is 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 clear 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 well-developed theories. That may be true. But such descriptions of the intonation of this language are actually driven by our own *impromptu* 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 independent assessment as to their validity.

Those who want to avoid the problem of having to invent their own theories when studying spontaneous speech often turn to a "standard" or widely used labeling scheme. What they may not be fully aware is that labeling and analysis, if performed at the same time, constitute an inherently circular process, as noted by Beckman (1997:12). That is, the labeling procedure assumes that we already know what and how to label, but the analysis procedure 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 instrumental observations. For example, in the ToBI conventions of labeling intonation, pitch accents labels are attached to the visually prominent F_0 peaks and valleys. An analysis of the corpus based on these labels is therefore taking for granted the assumptions behind those 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 Wightman (2002:28) has suggested, i.e., to "label what you hear", we are still left with the assumption that the labelers *know what to listen for* in the uncontrolled speech utterances.

3.3. The uses of spontaneous speech

There is no denying that spontaneous speech corpora are valuable for many purposes. They may be useful in motivating new hypotheses and raising questions about existing ones. On the other hand, theoretical postulations do not necessarily have to be based on direct observations. This is because how a theory is initially conceived is irrelevant to science according to the Popperian view (Popper, 1959). For speech, and especially for speech prosody, many existing theories are proposed largely based on introspection and nonsystematic observations. What is critical is that theories, regardless of how they are initially conceived, need to be tested through falsification. Spontaneous corpora can also provide us with frequency counts, such as how many yes-no questions are said with rising intonation and how many declarative statements are spoken with falling intonation (Hedberg, 2004). Also, as mentioned before, the validity of findings based on lab speech can be checked against spontaneous speech, as done in Bruce and Touati (1992). Nevertheless, it should be noted that for the spontaneous speech samples, identifying all the contributing factors is always tricky, as mentioned in 3.1. Thus great care and ingenuity are needed to make the validations credible.

Yet another potential use of spontaneous speech is related to a theoretical issue about speech acquisition. When children acquire a language, they need to learn everything that is critical for conducting successful communication in the language. But they have to achieve this feat without the benefit of controlled experimentation. Thus the mechanism of knowledge attainment in speech acquisition must be different from that in scientific inquiries. It could be the case that speech acquisition involves processes that are more akin to algorithms that can handle sparse data (van Santen, 1993) or rare events (Möbius, 2003). However, speech acquisition differs from these algorithms in that it is more directly guided by communicational functions. More research is needed on this important issue.

Finally, there are cases in which spontaneous utterances are the only speech samples one has access to, for example in the case of some endangered languages for which controlled experiments are rather difficult. But even in those cases, it is still important to apply the control principles as much as possible, because the level of certainty about any pattern in a language would be commensurate with the confidence one has over the consequence of deliberate manipulations related to that pattern.

4. Non-lexical communicative functions

The foregoing discussion suggests that much of the controversy over lab speech is driven by the need to better understand non-lexical functions in speech. Because most of these functions, unlike lexical contrast, are not represented orthographically (except for a limited number of punctuations), their control in the laboratory is not straightforward. As argued in the last section, looking at spontaneous speech directly may not be the right way to resolve this difficulty. This is because our research goal is not so much to ensure that the speech samples to be examined are rich enough to have a high probability of containing various communicative functions, but to clearly identify specific functions and understand how they are phonetically encoded.

The key to identifying non-lexical functions, I believe, is to start with the assumption that to be able to operate effectively, they must be sufficiently independent of each other, just like the lexical contrast function must have been sufficiently independent of other functions, for otherwise it would not have been possible for us to identify the basic acoustic properties of

vowels and consonants given the widely recognized lack of control for other functions in early research. Based on this assumption, we should also be able to, at least in theory, identify non-lexical functions one at a time. In practice, however, it is crucial to find strategies that are effective in revealing individual functions and identifying their mechanisms of encoding. For this purpose I have proposed a set of principles that may facilitate the identification of communicative functions (Xu, 2006):

1. *Specificity*. Each proposed communicative function should be as specific as possible in terms of the contrast it makes and in terms of the temporal domain of its operation.
2. *Mutual-exclusivity*. Each function should have a unique "encoding scheme" which has at least one predominant characteristic not overlapped by other functions. Thus once an observed pattern has been attributed to a particular function, it should not be reattributed to another function, unless there is clear evidence that they can both remain operative despite the overlap.
3. *Audibility*. A functional contrast in a language must have reached certain perceptual threshold, otherwise it would not have been operational. Exactly what the threshold is like (e.g., just above chance or much higher?), however, is itself a question that needs to be investigated in further research.
4. *Elicitability*. For a function to be verifiable, there needs to be at least one way of reliably eliciting it under experimental conditions. An unelicitable function is an unproven function.

The advantage of these principles is that they allow cumulative development of knowledge about communicative functions. This is seen in the fact that studies whose designs are consistent with these principles in one way or the other have been able to show evidence of distinct encoding schemes for focus (Cooper et al., 1985; Xu, 1999; Xu & Xu, 2005), sentence type (Eady & Cooper, 1986; Liu & Xu, 2005, 2007), new topic (Lehiste, 1975; Wang & Xu, 2006) and grouping/demarcation (Turk & Shattuck-Hufnagel, 2000; Wagner, 2005; Xu & Wang, 2009). Such evidence in turn demonstrates the likelihood of these being independent communicative functions. In contrast, however, we have yet to see how it is possible to apply these principles in spontaneous speech corpora, or how it is possible for research based on spontaneous speech alone to demonstrate evidence of communicative functions with equal clarity and consistency.

5. Lessons from other disciplines

Finally, it should 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 (1989:1189):

> … 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.

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. (p. 1185)

It may be too early to jump to conclusions about the fruitfulness of research based on spontaneous speech. But it is interesting to take a note of a report to the US Office of Naval Research by Garner (1950, cited by Banaji and Crowder 1991:78) comparing the relative worth of laboratory experimentation versus experimental manipulations in the operational field, which concludes that: "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 the two methods being compared are both experimental, but the one that is more "ecologically valid", i.e., the operational method, turned out to be more costly and less likely to produce generalizable data. In the case of speech research, a similar question may be asked: which is more cost effective, to maximize ecological validity or to maximize the level of control?

6. Conclusion

Despite its increasing unpopularity, many of us are still using lab speech in our research on both the lexical and non-lexical 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 hope to 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 devoid of communicative functions, interactions and emotions. 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 a fundamental limitation 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 studied also in the laboratory.

I have also argued that although spontaneous speech corpora no doubt have many uses, 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" (Banaji & Crowder, 1989:1192; Popper, 1959). 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 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. But marginalizing lab speech is clearly the wrong way to go.

Acknowledgement

I thank two anonymous reviewers for their comments on an earlier version of the paper. A previous, shorter version of this paper was presented at the 8th Phonetic Conference of China and appeared in Festschrift for Professor Wu Zongji's 100th birthday, 2009.

References

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.

Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *The Journal of the Acoustical Society of America*, *126***,** 2649-2659.

Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist*, *44***,** 1185-1193.

Banaji, M. R., & Crowder, R. G. (1991). Some everyday thoughts on ecologically valid methods. *American Psychologist*, *46***,** 78-79.

Barry, W. J. (1981). Prosodic functions revisited again! *Phonetica*, *38***,** 120-134.

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

Beckman, M. E. (1997) A typology of spontaneous speech. In *Computing Prosody: Computational Models for Processing Spontaneous Speech* (Y. Sagisaka, N. Campbell, & N. Higuchi, editors), pp. 7–26. New York: Springer Verlag.

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.

Bruce, G., & Touati, P. (1992). On the analysis of prosody in spontaneous speech with exemplification from Swedish and French. *Speech Communication*, *11***,** 453-458.

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.

Cheng, C., & Xu, Y. (2009) Extreme reductions: Contraction of disyllables into monosyllables in Taiwan Mandarin. In *Proceedings of Interspeech 2009*, Brighton, UK**,** pp. 456-459.

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.

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.

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.

Edlund, J., & Heldner, M. (2005). Exploring Prosody in Interaction Control. *Phonetica*, *62***,** 215-226.

Evans, B. G., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America*, *115***,** 352-361.

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America*, *121***,** 3814-3826.

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.

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, *36*, 680-703.

Gagne, J.-P., Rochette, A.-J., & Charest, M. (2002). Auditory, visual and audiovisual clear speech. *Speech Communication*, *37*, 213-230.

Gandour, J. (1999). Effects of speaking rate on Thai tones. *Phonetica*, *56*, 123-134.

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.

Garofolo, J., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L., & Zue, V. (1993) *TIMIT Acoustic-Phonetic Continuous Speech Corpus*. Philadelphia: Linguistic Data Consortium.

Gay, T. J. (1968). Effect of speaking rate on diphthong formant movements. *Journal of the Acoustical Society of America*, *44*, 1570-1573.

Gay, T. J. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, *63*, 223-230.

Gussenhoven, C. (2004) *The Phonology of Tone and Intonation*: Cambridge University Press.

Harnsberger, J. D., Wright, R., & Pisoni, D. B. (2008). A new method for eliciting three speaking styles in the laboratory. *Speech Communication*, *50*, 323–336.

Hedberg, N., Sosa, J. M., & Fadden, L. (2004) Meanings and configurations of questions in English. In *Proceedings of International Conference on Speech Prosody 2004*, Nara, Japan, pp. 309-312.

Hegde, M. N. (1994). Clinical research in communicative disorders: Principles and strategies. Austin, TX: PRO-Ed.

Hirata, Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*, *32*, 565-589.

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

Janse, E. (2004). Word perception in fast speech: artificially time-compressed vs. naturally produced fast speech. *Speech Communication*, *42*, 155-173.

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

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.

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 & T. Riad, editors), pp. 211-237. Berlin: Mouton de Gruyter.

Ladd, D. R. (2008) *Intonational phonology*. (Second ed.). Cambridge: Cambridge University Press.

LDC (2009) *Linguistic Data Consortium Home Page*. http://www.ldc.upenn.edu/

Lehiste, I. (1975) The phonetic structure of paragraphs. In *Structure and process in speech perception* (A. Cohen & S. E. G. Nooteboom, editors), pp. 195-206. New York: Springer-Verlag.

Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica*, *62*, 70-87.

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

Miller, J. L., O'Rourke, T. B., & Volaitis, L. E. (1997). Internal structure of phonetic categories: Effects of speaking rate. *Phonetica*, *54***,** 121-137.

Möbius, B. (2003). Rare events and closed domains: Two delicate concepts in speech synthesis. *International Journal of Speech Technology*, *6***,** 57-71.

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.

Nakajima, S., & Allen, J. F. (1993). A study on prosody and discourse structure in cooperative dialogues. *Phonetica*, *50***,** 197-210.

Ostendorf, M., Price, P., & Shattuck-Hufnagel, S. (1996) *Boston University Radio Speech Corpus*. Philadelphia: Linguistic Data Consortium.

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.

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.

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.

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.

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.

Popper, K. (1959) *The Logic of Scientific Discovery (translation of Logik der Forschung)*. London: Hutchinson.

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***,** 473-500.

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.

Rischel, J. (1992). Formal linguistics and real speech. *Speech Communication*, *11***,** 379-392.

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.

Shih, C. (1993) Relative prominence of tonal targets. In *Proceedings of The 5th North American Conference on Chinese Lingusitics*, Newark, Delaware**,** pp. 36.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40***,** 227-256.

Schlosser, R. W. (2003) The Efficacy of Augmentative and Alternative Communication: Toward Evidence-Based Practice: Toward Evidence-based Practice (Augmentative and Alternative Communications Perspectives) San Diego: Lyle L. Lloyd.

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

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. In *Proceedings of The 1992 International Conference on Spoken Language Processing*, Banff**,** pp. 867-870.

Tiffany, W. R. (1980). The effects of syllable structure on diadochokinetic and reading rates. *Journal of Speech and Hearing Research*, *23*, 894-908.

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, *28*, 397-440.

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. Nooteboom* (H. Quené & J. V. v. Heuven, editors), pp. 83-93. The Netherlands: LOT Occasional series by Utrecht University.

Van Santen, J. P. H. (1993). Exploring N-way Tables with Sums-of-Products Models. *Journal of Mathematical Psychology*, *37*, 327-371.

Wagner, M. (2005). *Prosody and Recursion*. Ph.D. Dissertation, Massachusetts Institute of Techonology.

Wang, B., & Xu, Y. (2006) Prosodic encoding of topic and focus in Mandarin. In *Proceedings of Speech Prosody 2006*, Dresden, Germany, pp. PS3-12_0172.

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.

Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, *18*, 3-35.

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.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, *95*, 2240-2253.

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, *55*, 179-203.

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

Xu, Y. (2001). Fundamental frequency peak delay in Mandarin. *Phonetica*, *58*, 26-52.

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

Xu, Y. (2006) Speech prosody as articulated communicative functions. In *Proceedings of Speech Prosody 2006*, Dresden, Germany, pp. SPS5-4-218.

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.

Xu, Y. (2009). Timing and coordination in tone and intonation--An articulatory-functional perspective. *Lingua*, *119*, 906-927.

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.

Xu, Y., & Wang, M. (2009). Organizing syllables into groups—Evidence from F_0 and duration patterns in Mandarin. *Journal of Phonetics*, *37*, 502-520.

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

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.

---

[i] A caveat is that in some cultures written texts are always associated with a formal style of speaking, or represent only the standard dialect. In that case, it may require greater ingenuity in the research design to elicit the desired style of spoken form.