Information Flow and Plasticity across Levels of Linguistic Sound Structure: Responses to the Target Papers by Cutler & Weber and by Goldinger.

Ann Bradlow

Northwestern University
abradlow@northwestern.edu

ABSTRACT

The two target papers in this session present different approaches to substantiating the consensus position that both abstract and episodic levels of representation underlie speech processing. Cutler and Weber's paper provides an interesting illustration of the non-direct phonetic-to-lexical mapping in second language speech processing. Their contribution presents a series of studies in which meta-knowledge of a second language contrast facilitated representation of the contrast at a lower level. As a complement to this contribution, I discuss cases in which the *lack* of access to abstract category representations effectively blocks the passage of information across the episodic-to-abstract hierarchy. Second, picking up on the challenge of specifying the nature of exemplars as discussed in Goldinger's target article, I discuss recent data regarding experience-dependent shaping of the basic sensory circuitry. These data are particularly relevant to the case of second language speech learning as they help specify the levels of plasticity required for successful non-native language acquisition.

Keywords: second language, speech perception, exemplar models

1. Phonetic-to-Lexical Scaling-Up Problem

Cutler and Weber's target paper shows how meta-knowledge regarding the existence of a lexical contrast in a non-native language "makes its way into" online word recognition even though the critical phoneme contrast may be very poorly perceived by non-native listeners. This important demonstration is consistent with several other demonstrations of the "scaling up" problem, that is, the problem that patterns of phoneme discrimination and identification are very often not directly reflected in patterns of word or sentence recognition. Cutler and Weber offer a partial solution to this problem based on meta-knowledge that is available to a second-language learner (but not to an infant during first language acquisition). In particular, the case in their report is one in which apparent neutralization of a phoneme contrast does not prevent sensitivity to the relevant lexical contrast thanks to meta-knowledge of the contrast from orthography. In this section I expand on this contribution by discussing cases of the "scaling up" problem in non-native speech processing in which the overall result is unfortunately less positive for the second language learner. In these cases, fairly accurate processing at a lower level of representation, where access to the full-blown hierarchy of category labels is irrelevant, does not fully determine processing accuracy when more complex input is encountered. Whereas the Cutler and Weber case involves contrast neutralization at the lower, phonetic level and contrast representation at the higher, orthographic level, the cases I discuss below involve native like patterns at a lower level of phonetic processing but poorer-than-native performance at a higher level of phonetic processing.

The first case of nonlinear scaling up is the case of non-native speech-in-noise perception where we observe a non-uniform nonnative deficit as we move from phoneme identification up to sentence recognition. Several studies have shown that, despite native-like levels of word and sentence recognition accuracy in listening conditions with no background noise or reverberation, highly proficient non-native listeners typically show sharper declines in word and sentence recognition accuracy than their native listener counterparts [e.g. 6, 7, and 9]. This finding is, of course, perfectly consistent with the common experience of individuals listening to a non-native language; somehow the volume control on a radio always needs to be adjusted as you switch from listening to speech in your native language to listening to speech in a non-native language. The fact that even highly proficient bilinguals experience greater difficulty with speech-in-noise perception than monolinguals has important practical implications for bilinguals in educational and occupational settings, and there is clearly a need to more fully understand the causes and consequences of this basic phenomenon.

One possible explanation for this phenomenon is that the source of the trouble for non-native listeners is at the level of phoneme perception. Under this possibility we expect to also see disproportionate declines in phoneme-in-noise perception with decreasing signal-to-noise ratios by non-native listeners relative to their native-listener counterparts. However, recent data [4 and 5] indicate a more complicated phoneme-to-word/sentence mapping. Specifically, parallel effects of increasing noise (decreasing signal-to-noise ratio) for native and non-native listeners (Dutch natives) were found in a task of English phoneme-in-noise perception (using VC and CV stimuli) [4]. If phoneme perception fully determined word and sentences perception, then we would expect the observed parallel effects of increasing noise levels for native and non-native listeners at the phoneme level to be reflected throughout the sound structure hierarchy (from phoneme to word to sentence). However, this does not appear to be the case; instead, the constant non-native deficit in phoneme-in-noise perception across signal-to-noise ratios appears to "take a big hit" once access to lexical- and phrase-level information comes into play. In other words, there is a strong effect of the accumulation of native-to-target language mismatches across levels of processing.

An alternative means of examining noise-related declines for native and non-native listeners is to examine the effects of different noise types on speech recognition accuracy. Since speech-in-noise perception is subject to a combination of energetic and informational masking of the target speech by the background noise, different types of noise exert their masking influences at different levels. That is, the ratio of informational to energetic masking decreases from a single-talker distracter to multi-talker babble to broadband noise. Accordingly, a recent study ([5]) compared consonant identification accuracy (in VCV frames) in various types of noise at a fixed signal-to-noise ratio by native (English) and non-native (Spanish) listeners. The results showed that native listener consonant identification accuracy declined across the various conditions going from quiet listening conditions to speech with a single-talker distracter to speech embedded in speech-shaped noise and finally, to speech in multi-talker babble. Interestingly, non-native listeners followed this same order of declining consonant recognition accuracy, and in this case, the non-native listener decline was generally sharper than the native listener decline across these conditions. However, an important difference between the native and non-native listeners emerged in the single-talker distracter condition. While the native listeners performed better in non-native language distracter noise (Spanish) than in the native language distracter noise (English) [see also 10], the non-native listeners showed no language-of-noise difference. The net effect of this difference was a greater non-native listener deficit in Spanish noise than in English noise. Thus, despite equivalent changes in lower-level, energetic masking, the higher-level influence of the speech masker exerted an effect that depended on the listener's level of experience with the masker and signal languages. Familiarity with both the masker and signal languages (as was the case for the nonnative listeners in this experiment) resulted in a fairly symmetrical pattern of phoneme identification accuracy across the two noise-language conditions. In contrast, familiarity with just the target language (as was the case for the native listeners) resulted in an asymmetric pattern of phoneme identification accuracy across these conditions. This overall pattern suggests a complex interaction of signal-dependent (energetic masking) and signal-independent (informational masking) processes that act upon the input in qualitatively similar yet quantitatively different ways for native and nonnative listeners.

Both of these studies are imperfect tests of the directness of the phoneme-to-word/phrase mapping. In neither study were the non-native listeners selected on the basis of native-like levels of performance in the most favorable conditions, which suggests that they came from a somewhat different non-native population to the highly proficient listeners in the studies of word- and sentence-in-noise perception [6, 7, and 9]. (However, note that there was a positive correlation between amount of performance degradation and consonant identification rate in noise for the non-native listeners in [5]). Clearly more extensive and better controlled comparisons across native and highly proficient non-native listeners on phoneme. word and phrase recognition in various noises types and levels are needed. Nevertheless, this general research area provides further illustrations of the indirect and asymmetrical phonetic-to-lexical mapping discussed by Cutler and Weber.

A second case of phonetic-to-lexical asymmetry is the case of Mandarin lexical tone discrimination by native and non-native (American English) listeners with no tone language experience [2]. In this study, discrimination of the Mandarin four-way lexical tone contrast was compared across native and non-native listeners in monosyllabic and tri-syllabic utterances. The tri-syllabic utterances consisted of the target monosyllable preceded and followed by a syllable with the same phonemes as the target syllable (namely, 'ra-ra-ra'). All comparisons were within a constant tri-syllabic frame, in which only the tone on the middle syllable varied. Both groups of listeners were highly sensitive (d' > 3.5) to all of the tonal contrasts in monosyllables, indicating that both groups had quite accurate, phonetic level encoding of the stimulus. The Mandarin listeners also showed high sensitivity to almost all of the tonal contrasts when they occurred on the middle syllable of a tri-syllabic frame. In contrast, the English listeners sensitivity to the tonal contrasts in the tri-syllabic frames was highly dependent on the physical, acoustic similarity between the stimuli: the closer the stimuli in terms of their F0 contours, the lower the English listeners' sensitivity to the tonal contrast. Thus, assuming that the episodic encoding of these F0 contours was quite veridical for the English listeners, the absence of the category contrast at the abstract level forced strong reliance on the more episodic level of encoding. For these listeners, unlike the more fortunate listeners in the case discussed by Cutler and Weber, the phonetic encoding of the stimuli had no further support from knowledge of the existence of the contrast at a more abstract level. In contrast, for the Mandarin listeners, the existence of abstract, linguistic category contrasts facilitated processing in the discrimination task; in cases of a high degree of acoustic similarity, the categorical contrast served to reinforce any fragility of the acoustic/episodic-level encoding.

Taken together, these cases indicate that the indirect mapping between lower and higher levels of processing can result in various patterns of perception for native and non-native perception. The general point then, echoing Cutler and Weber's conclusion, is that second language speech learning is all about gaining the knowledge structures and attentional weights that support a sufficiently native-like set of associations across the episodic-to-abstract hierarchy. Moreover, the acquisition of the necessary knowledge structures and attentional weights can occur independently, through means that are not necessarily identical to those of first language acquisition, and can interact in ways that can either aid or impede second language processing.

2. Experience-Dependent Shaping of the Basic Sensory Circuitry

Goldinger's target paper develops in some detail the notion of complementary episodic and abstract systems. In applying this kind of neurologically-plausible model to the case of second language acquisition, an important question to ask is: To what extent are the very early stages of auditory signal encoding subject to experience-dependent learning? Goldinger points out that a "rational" theory must admit an effect of psychological experience on memories of experienced events such that "...each stored 'exemplar' is actually a product of perceptual input combined with prior knowledge." Taking this a step further, I discuss some recent behavioral and neurophysiologic evidence for an influence of long-term experience on auditory encoding at surprisingly "basic" levels [1, 3, 11]. This then supports the suggestion that high proficiency in a non-native language must involve modifications throughout the episodic-to-abstract hierarchy.

One strategy for investigating the influence of long-term linguistic experience across levels of processing is to compare both speech and non-speech perception across listeners from different native language backgrounds. If native language experience exerts an influence on non-speech perception as well as on speech perception then the influence of linguistic experience would appear to be pervasive and could possibly extend to the very earliest stages of episodic encoding of auditory experiences. Following this general reasoning, a recent study tested Mandarin and English listeners on a set of auditory perceptual tasks ranging from pitch discrimination in simple, non-speech signals, to identification of non-speech pitch glides as level, falling or rising, to Mandarin lexical tone identification [3]. Not surprisingly, there was a strong effect of language background on the lexical tone identification task: the Mandarin listeners identified Mandarin tones significantly more accurately than the English listeners. Performance did not differ across the Mandarin and English listeners on the pitch discrimination task with simple, non-speech sounds suggesting that their basic pitch discrimination abilities did not differ significantly. However, cross-language differences were observed on the non-speech pitch contour identification task: While the English listeners were quite accurate at identifying the direction of pitch changes in non-speech pitch glides, the Mandarin listeners often misidentified flat and falling pitch glides in a manner that reflected features of Mandarin sound structure. This finding suggests that the influence of native language learning can extend beyond the perception of speech signals to the encoding and processing of non-speech, auditory input.

A similar conclusion can be drawn from a separate study in which long term auditory experience was manipulated by comparing non-native language speech perception by musicians and non-musicians [1]. In this study, English listeners with extensive musical experience showed an advantage over their non-musician counterparts in Mandarin tone identification and discrimination. In a second experiment, identification and discrimination of brief musical phrases (sequences of 5 tones) was compared across native speakers of a tone language (Mandarin) versus a non-tone (English) language. In this experiment, the Mandarin listeners more accurately discriminated, but less accurately identified, the short musical melodies relative to the English listeners. These data suggest that, in low-level discrimination tasks, long-term experience with pitch in one domain (through either native language experience or extensive musical training) generalized to the other domain (language or music) leading to enhanced pitch discrimination in that domain. But in high-level identification tasks, long-term experience with pitch categories in one domain interfered with pitch category labeling in the other. The purpose of this study was to investigate the relationship between pitch processing in music and speech; however, for our purposes, it serves to make the point that long-term auditory experience that may be confined to one domain (e.g. music or speech) shapes perception in a very general way.

The patterns observed in these behavioral studies have recently been corroborated by neurophysiologic evidence [11]. Specifically, a comparison of F0 encoding at the auditory brainstem across musicians and non-musicians showed higher fidelity and greater robustness of F0 encoding for the listeners with extensive musical training. This then suggests that long-term auditory experience may have an influence on the earliest, subcortical levels of auditory signal encoding and therefore that the very basic sensory circuitry is shaped by experience-dependent learning. It is important to note that, within the group of musicians, fidelity and robustness of F0 encoding at the level of the auditory brainstem was positively correlated with the number of years of musical training and negatively correlated with the age of onset of musical training. This indicates that the observed group-wise, musician-nonmusician F0 brainstem encoding difference was unlikely to be due to genetic differences between the groups, but was instead due to differences in auditory experience.

Extending this general line of reasoning to the second language speech learning situation suggests both good and bad news. On the one hand, all levels of auditory encoding and processing -- from the most basic, sensory encoding that underlies episodic traces of auditory input to the most abstract level of linguistic, categorical structure -- are likely to be tuned to the native language sound structure and therefore require modification for accurate second language acquisition. On the other hand, it is also likely that the entire system (from episodic encoding to abstract category label establishment) is under continuous updating and remains highly plastic in response to current input throughout the life-span.

CONCLUDING REMARKS

Taken together, the target papers of this special session set the stage for a potential breakthrough in theories of second language speech learning. Goldinger's contribution demonstrates the theoretical progress that has been made over the past 10-15 years [see also 8] as well as the current, neurologically-oriented direction of further model development. The contribution by Cutler and Weber illustrates the sophisticated empirical base that we can now build due to theoretical, technical and subject sampling advances. As I hope to have shown in this commentary, there is a great need for more empirical studies that provide well-controlled comparisons across tasks that probe processing and representation at various levels in populations with various input experiences. This then may lead to a detailed, formal and insightful exemplar based theory of second language speech learning.

ACKNOWLEDGEMENT

Work supported by NIH Grant R01 DC005794.

REFERENCES

[1] Alexander, J. A., Bradlow, A. R., Ashley, R. D. & Wong, P. C. M. (manuscript under review) On the interaction of speech-pitch processing and music-pitch processing.

[2] Bent, T. 2005. Perception and production on non-native prosodic categories. PhD dissertation. Northwestern Univ.

[3] Bent, T., Bradlow, A. R., & Wright, B. A 2006. The influence of linguistic experience on pitch perception in speech and non-speech sounds. *JEP:HPP*, 32(1), 97-103

[4] Cutler, A., Weber, A., Smits, R. & Cooper, N. 2004. Patterns of English phoneme confusion by native and non-native listeners. *J. Acoust. Soc. Am.*, 116(6), 3668-3678.

[5] Garcia Lecumberri, M. L. & Cooke, M. 2006. Effect of masker type on native and non-native consonant perception in noise. *J. Acoust. Soc. Am.*, 119(4), 2445-2454.

[6] Mayo, L. H., Florentine, M., & Buus, S. 1997. Age of Second-Language Acquisition and Perception of Speech in Noise. *J. Speech Lang. Hear. Res*. 40, 686-693.

[7] Meador, D., Flege, J. E. & MacKay, I. R. 2000. Factors affecting the recognition of words in a second language. *Bilingualism* 3, 55-67.

[8] Pierrehumbert, J. 2006. The next toolkit. *J. Phonetics*, 34(6), 516-530.1.

[9] Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M. & Abrams, H. B. 2006. Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics* 27, 465-485

[10] Van Engen, K. J. and Bradlow, A. R. 2007. Sentence recognition in native- and foreign-language multi-talker background noise. *J. Acoust. Soc. Am.*, 121(1), 519-526.

[11] Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T. & Kraus, N. 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10(4), 420-422.