# PROCESSING RELATIONSHIPS BETWEEN LANGUAGE-BEING-SPOKEN AND OTHER SPEECH DIMENSIONS

Charlotte Vaughn[1] & Ann R. Bradlow[2]

University of Oregon, USA[1], Northwestern University, USA[2]
cvaughn@uoregon.edu[1], abradlow@northwestern.edu[2]

## ABSTRACT

While indexical information is implicated in language processing, little is known about the internal structure of the system of indexical dimensions itself, particularly in bilinguals. A series of three experiments using the speeded classification paradigm investigated the relationship between various indexical and non-linguistic dimensions of speech in processing (talker identity, talker gender, and amplitude of speech) and a lesser-studied indexical dimension relevant to bilinguals, namely, which language is being spoken. Results demonstrate that language-being-spoken is integrated in some form with each of the other dimensions tested, and that this relationship is independent of listeners' bilingual status.

**Keywords**: speech perception, selective attention, bilingualism, indexical information, Garner task.

## 1. INTRODUCTION

Research has increasingly shown that indexical information is implicated in language processing [12, 16, 18]. However, little is known about the internal structure of the system of indexical dimensions (e.g., talker identity, gender, age, the language being spoken, etc.) itself. Due to the complex and multidimensional nature of the speech signal, individual dimensions are rarely, if ever, encountered on their own. Thus, knowledge of the processing relationships between different types of indexical and other non-linguistic information is important to understanding human language processing in general.

Even less is known about relationships between indexical dimensions in bilinguals, which may be different than in monolinguals for several reasons. Bilinguals must make associations between talkers and the languages those talkers speak, meaning that they may have more experience than monolinguals in attending to indexical information, and, those associations may be useful in language segmentation for simultaneous bilinguals [26]. Also, bilinguals have shown advantages over monolinguals in tasks involving executive control [1, 2], a benefit that may extend to the domain of indexical processing.

Talker and language information seem to be associated in processing, as demonstrated by the language familiarity benefit, where listeners identify talkers better in familiar languages than in unfamiliar languages [6, 11, 14, 20, 27]. While these results indicate that talker and language information are related in processing, there has been no direct test to determine whether they recruit the same processing resources, and whether the processing relationship between the two dimensions is symmetrical or asymmetrical.

Moreover, it is not known whether there is a hierarchy of processing relationships between speech dimensions, indexical (e.g. gender of talker) or otherwise (e.g. amplitude of speech). This is a notable gap in the literature given that a good deal is known about the processing relationships between indexical and linguistic dimensions (such as the identity of a segment in a syllable or word [9, 16]).

Thus, the present paper reports on a series of experiments that employed the speeded classification paradigm (or Garner task [10]) to explicitly examine the relationships between various indexical and non-linguistic dimensions of speech in processing. In this task, listeners make a classification decision within one dimension (e.g. is the talker male or female?) while irrelevant variation in another dimension (e.g. the language of the stimuli, Chinese or English) is either present or absent in that block of stimuli. A processing dependency, also called *interference*, between two dimensions is indicated by greater reaction times to stimuli in blocks with variability from another dimension (e.g. identifying talker gender when the language of the speech also varies) than to stimuli in blocks with no irrelevant variability (e.g. identifying talker gender when the language is constant).

The present studies, which tested both monolingual and bilingual participants, investigated processing relationships between which language is being spoken (e.g. *language-being-spoken*, or *L-B-S*; in these experiments Mandarin Chinese or English) and three other speech dimensions (*talker* identity in Experiment 1[1], talker *gender* in Experiment 2[2], and *amplitude* of speech in Experiment 3). This set of experiments compared L-B-S against three dimensions chosen to differ from L-B-S in several key ways that may affect processing. The processing of L-B-S and talker (Expt 1) are similar in that: both processes require the use of multiple acoustic cues ([13, 25] for talker, [17, 22] for L-B-S), both constitute choosing from members of an open set (i.e. the set of all human languages, and of all talkers, though bounded to a binary decision in these experiments), and both dimensions are used by listeners in the course of speech comprehension. Unlike L-B-S, amplitude (Expt 3): is a single acoustic cue, is not an indexical property, its categorization as "loud" or "soft" does not require a priori knowledge, and its processing is more peripheral to speech comprehension [5, 19]. Similar to L-B-S processing, gender processing (Expt 2): does not rely solely on one acoustic cue (though F0 is dominant [7, 8]), but unlike L-B-S, gender is a closed set (male or female).

Results of these experiments will indicate (a) to what extent different speech dimensions recruit similar resources in processing, indicative of how interconnected the system of indexical dimensions is, and (b) whether this cognitive architecture is similar for bilinguals and monolinguals.

## 2. METHODS

Table 1 summarizes the speech dimensions tested in each experiment, the talkers who produced the stimuli, and the language background of the participants.

### 2.1. Stimulus materials and talkers

Stimuli were taken from the ALLSSTAR corpus [4], and were short, meaningful sentences originally developed for the Hearing in Noise Test (HINT, English version [21]; Mandarin version [28]), e.g. "Somebody stole the money". Sixty-four sentences read in English and 64 read in Mandarin were selected for use in all experiments. Stimuli in Experiments 1 and 2 were amplitude-normalized to 70 dB SPL, a comfortable listening level. In Experiment 3, because of the explicit amplitude manipulation, the stimuli were amplitude-normalized to 45 dB SPL for "soft" and 75 dB SPL for "loud" stimuli.

The stimulus talkers were late bilinguals, L1 Mandarin Chinese and L2 English, all in their early 20s, and who had all recently moved to the greater Chicago area.

### 2.2. Participants

In Experiment 1, participants from three language backgrounds were tested (18 participants from each group), all of whom were recruited from the undergraduate population of Northwestern University or the greater Chicago area. The ENG group consisted of English monolinguals who did not know Mandarin Chinese; the MAN group were Mandarin-English bilinguals who were L2 English, L1-dominant; and the NMB group were bilinguals fluent in English and a language that was not Mandarin, and were L2 English, L1-dominant. Thus, the NMB group matched the MAN group in bilingual status, and the NMB group matched the ENG group in that both were only familiar with one of the languages being tested (English). The NMB group was included to assess the possibility that bilinguals in general may show different processing relationships between dimensions, regardless of their familiarity with the languages being tested. Experiments 2 and 3 each tested 18 different ENG participants and 18 different MAN participants.

**Table 1**: Summary of stimulus dimensions, talkers, and participants in Experiments 1-3.

| Expt | Stimulus dimensions | Chinese-English bilingual talkers | Participant groups by language background |
|---|---|---|---|
| 1 | **L-B-S** (Chinese/English) vs. **Talker** (Wei/Li) | 2 males | ENG (N = 18)<br>MAN (N = 18)<br>NMB (N = 18) |
| 2 | **L-B-S** (Chinese/English) vs. **Gender** (male/female) | 1 male & 1 female | ENG (N = 18)<br>MAN (N = 18) |
| 3 | **L-B-S** (Chinese/English) vs. **Amplitude** (loud/soft) | 1 male | ENG (N = 18)<br>MAN (N = 18) |

### 2.3. The speeded classification task

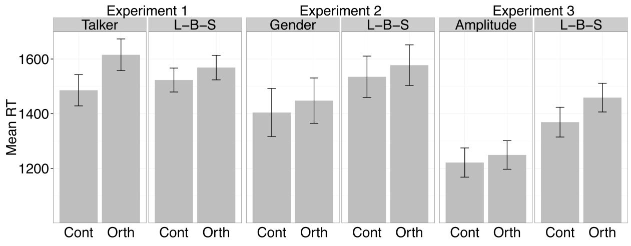
Participants in each experiment completed six blocks of the speeded classification task, corresponding to three different stimulus sets (orthogonal, control, and correlated[3]) for two dimensions of the stimulus per experiment (e.g., in Experiment 2, gender and language-being-spoken). In the orthogonal blocks, the two dimensions varied independently (e.g. Chinese sentences were presented in both the male and female voice, and English sentences were also presented in both the male and female voice). Thus, in orthogonal blocks classification along one dimension requires ignoring variation in the other, irrelevant dimension. In the control blocks there is no irrelevant variation (e.g. only Chinese sentences are presented for gender classification). The dependent variable is the difference in reaction time (RT) between orthogonal and control blocks. If listeners can ignore irrelevant variability in one dimension when classifying the other, they will take no longer to perform the classification in orthogonal than in control blocks, and those dimensions are said to be *separable*. If, however, listeners cannot ignore irrelevant variation in one dimension when classifying the other, their RTs in orthogonal blocks will be longer than in control blocks, and the two dimensions are said to be *integral*. Further, dimensions may not always show symmetrical integration; if one task shows more interference than the other the relationship between those two dimensions is *asymmetrical*, suggesting that one dimension is more salient in processing.

The conditions were blocked by stimulus dimension, so that participants received three blocks to be classified by one dimension, and then the three blocks for the other dimension.

### 2.4. Procedure

Stimulus sentences were presented one at a time over headphones to participants who were seated in a sound attenuated booth. Participants were asked to classify each sentence according to the choices presented on the button box for the task at hand (e.g., “loud” or “soft” for the amplitude task of Expt 3), and were told that they did not need to wait until the sentence was over before responding. At the beginning of each task participants heard 16 practice trials with feedback. One additional instruction was given in the talker task of Experiment 1, since distinguishing between the talker named “Wei” and the talker named “Li” was the only classification decision for which listeners would have no a priori knowledge. Therefore in that task they were told to use trial and error to figure out which talker was which in the practice trials.

**Figure 1**: By-participant means of reaction times by task and block for all experiments (Cont = Control, Orth = Orthogonal). Error bars represent +/- 1 SE of the mean.



## 3. RESULTS

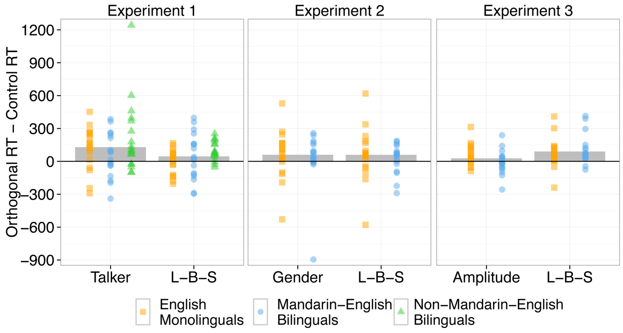
In all experiments, accuracy was near ceiling on all tasks, and therefore is not analyzed here. An analysis of reaction time is reported, using correct responses only. Outliers were trimmed by removing RTs > 2.5 standard deviations from each participant’s mean (by block, task, and experiment), resulting in the removal of ≤ 2.65% of a listener group’s RTs for each experiment. RTs were modeled using repeated measures ANOVAs on by-participant log-transformed means. The results are presented first collapsed across all participant language groups, and are visualized as bar plots in Fig. 1.

In Experiment 1, there was an integral yet asymmetrical pattern of interference between L-B-S and talker identity, where listeners were slower to respond in orthogonal (*M*=1592.07 ms) than control blocks (*M*=1504.47 ms), F(1,51)=24.437, p<0.0001, and this interference was greater in the talker task (129.64 ms difference between orthogonal and control) than in the L-B-S task (45.56 ms difference), F(1,51)=6.715, p=0.012. This indicates that it was harder for listeners to ignore L-B-S when attending to talker identity than the reverse.

In Experiment 2, there was an integral and symmetrical pattern of interference between L-B-S and gender. Participants were slower to respond in orthogonal (*M*=1512.57 ms) than control blocks (*M*=1469.50 ms), F(1,34)=5.647, p=0.023, but there was no significant interaction between task and block, F(1,34)=0.022, p=0.882.

In Experiment 3, L-B-S and amplitude showed asymmetrical interference, where listeners could not ignore amplitude when attending to L-B-S (89.7 ms difference), but could ignore L-B-S when attending to amplitude, F(1,34)=4.119, p=0.050.

**Figure 2**: Interference in all experiments. Grey bars represent group means of differences between orthogonal and control blocks (positive values indicate interference). Colored shapes represent interference levels by individual participant, split by language background.



In all three experiments, no effects of participant language background were found, as illustrated by the similar patterning across listener groups in Fig. 2. The amount of interference did not differ by participant language background for Expt 1 (F(2,51)=1.571, p=0.218), Expt 2 (F(1,34)=0.979, p=0.33), or Expt 3 (F(1,34)=0.406, p=0.528).

## 4. DISCUSSION

Language-being-spoken showed some form of integration with each of the other dimensions tested, indicating that it shares processing resources with those dimensions, and providing evidence that the internal system of indexical (and non-linguistic) speech dimensions is highly interconnected. However, the exact form of the processing relationship differed in each experiment: L-B-S interfered in talker processing more than vice versa (Expt 1), gender and L-B-S showed symmetrical interference (Expt 2), and amplitude interfered in L-B-S processing less than vice versa (Expt 3). Dimensions that are difficult to ignore (causing more interference) can be said to be more salient, as they capture more attention [23]. Framed in this way, the results can be schematized as a processing hierarchy of relative salience: language-being-spoken is as salient as gender, more salient than talker, and less salient than amplitude.

### 4.1. Talker processing is more reliant on language processing than vice versa

The results of Experiment 1 are aligned with previous literature on the role of language information in talker identification, and vice versa. In line with previous work demonstrating a language familiarity benefit for talker identification described in Section 1, it was found that L-B-S did interfere with the processing of talker. Further, the fact that talker also interfered with the processing of L-B-S is in line with previous work suggesting that listeners can use talker information when classifying languages [17, 22]. Additionally, the asymmetry found in this experiment, where there was more interference from language-being-spoken when processing talker than vice versa, is logical in light of the fact that talker-general cues may be more important than talker-specific cues in language identification; listeners abstract over talker information to form representations of languages, and phonetic information is organized language-specifically [3, 15]. Thus, it appears that language processing is not as reliant on talker processing as the reverse.

### 4.2. Bilinguals and monolinguals behave similarly

Finally, these results indicate that the processing relationships between these dimensions were the same for bilinguals and monolinguals. This finding is counter to hypotheses about differences based on bilingual advantages in executive control or based on bilinguals' increased focus on indexical information. Instead, listeners demonstrated similar processing relationships between speech dimensions regardless of their familiarity with the languages tested (since Mandarin-English bilinguals were not different than English monolinguals and non-Mandarin-English bilinguals), or their bilingual status (since English monolinguals were not different than both groups of bilinguals). Thus, the current results show that the integrality of L-B-S with other speech dimensions is independent of the language experience of listeners.

## 5. CONCLUSIONS

In sum, the processing of language-being-spoken is integrated in some way with the processing of talker, gender, and amplitude. Further, language familiarity and bilingual status do not affect language-being-spoken's relationship with these other dimensions in processing, suggesting that the cognitive architecture underlying relationships between speech dimensions in processing appears to be similar for bilinguals and monolinguals. In situating language-being-spoken with respect to other indexical and non-linguistic speech dimensions, this work contributes to the growing understanding of the position of indexical dimensions in human language processing.

## 6. ACKNOWLEGMENTS

Work on this project was conducted while the first author was at Northwestern University. The authors wish to acknowledge Susanne Brouwer, Vanessa Dopker, Chun Liang Chan, Tyler Kendall, Emily Kahn, and Lindsay Valentino. Work supported in part by Grant R01-DC005794 from NIH-NIDCD to the second author.

## 7. REFERENCES

[1] Bialystok, E., Craik, F. I. M., Klein, R., Viswanathan, M. 2004. Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*, 290–303.

[2] Bialystok, E., Martin, M. M. 2004. Attention and inhibition in bilingual children: Evidence from the developmental change card sort task. *Developmental Science, 7*, 325–339.

[3] Bradlow, A. R. 1996. A Perceptual Comparison of the /i/–/e/ and /u/–/o/ Contrasts in English and in Spanish: Universal and Language-Specific Aspects. *Phonetica*, *53*(1–2), 55–85.

[4] Bradlow, A. R., Ackerman, L., Burchfield, L. A., Hesterberg, L., Luque, J. S., Mok, K. 2010. ALLSSTAR: Archive of L1 and L2 Scripted and Spontaneous Transcripts And Recordings. Department of Linguistics, Northwestern University. Retrieved from http://groups.linguistics.northwestern.edu/speech_comm_group/allsstar/index.html.

[5] Bradlow, A. R., Nygaard, L. C., Pisoni, D. B. 1999. Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics, 61 (2)*, 206–219.

[6] Bregman, M. R., Creel, S. C. 2014. Gradient language dominance affects talker learning. *Cognition*, *130*(1), 85–95.

[7] Childers, D. G., Wu, K. 1991. Gender recognition from speech. Part II: Fine analysis. *J. Acoust. Soc. Am*. 90(4), 1841–1856.

[8] Coleman, R. O. 1971. Male and female voice quality and its relationship to vowel formant frequencies. *J. of Speech and Hearing Research, 14*, 565–577.

[9] Cutler, A., Andics, A., Fang, Z. 2011. Inter-dependent categorization of voices and segments. *Proc. 17th ICPhS* Hong Kong, 552–555.

[10] Garner, W. R. 1974. *The processing of information and structure*. Potomac, MD: Lawrence Erlbaum.

[11] Goggin, J. P., Thompson, C. P., Strube, G., Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, *19*(5), 448–458.

[12] Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological review*, *105*(2), 251–279.

[13] Klatt, D. H., Klatt, L. C. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am*. 87(2), 820–857.

[14] Köster, O., Schiller, N. O., Künzel, H. J. 1995. The influence of native-language background on speaker recognition. *Proc. 13th ICPhS* Stockholm, 306–309.

[15] Lisker, L., Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384–422.

[16] Mullennix, J. W., Pisoni, D. B. 1990. Stimulus variability and processing dependencies in speech perception. *Perception* & *Psychophysics, 47*, 379–390.

[17] Muthusamy, Y. K., Jain, N., Cole, R. A. 1994. Perceptual benchmarks for automatic language identification. *Proc. ICASSP-94.* Vol. 1, pp. I–333. IEEE.

[18] Nygaard, L. C., Sommers, M. S., Pisoni, D. B. 1994. Speech perception as a talker-contingent process. *Psychological Science, 5*, 42–46.

[19] Nygaard, L. C., Sommers, M. S., Pisoni, D. B. 1995. Effects of stimulus variability on perception and representation of spoken words in memory. *Attention, Perception, & Psychophysics*, *57*(7), 989–1001.

[20] Perrachione, T. K., Wong, P. C. M. 2007. Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia*, *45*(8), 1899–1910.

[21] Soli, S. D., Wong, L. L. N. 2008. Assessment of speech intelligibility in noise with the hearing in noise test. *Intl. J. Audiology* 47, 356–361.

[22] Stockmal, V., Muljani, D., Bond, Z. 1996. Perceptual features of unknown foreign languages as revealed by multi-dimensional scaling. *Proc. ICSLP,* Vol. 3, pp. 1748–1751. IEEE.

[23] Tong, Y., Francis, A. L., Gandour, J. T. 2008. Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*, *23*(5), 689–708.

[24] Vaughn, C. 2014. Language-Being-Spoken and Other Indexical Dimensions in Monolingual and Bilingual Speech Processing (Unpublished doctoral dissertation). Northwestern University, Evanston, IL, USA.

[25] Walden, B. E., Montgomery, A. A., Gibeily, G. J., Prosek, R. A., Schwartz, D. M. 1978. Correlates of psychological dimensions in talker similarity. *J. of Speech, Language, and Hearing Research*, *21*(2), 265–275.

[26] Weiss, D. J., Gerfen, C., Mitchel, A. D. 2009. Speech segmentation in a simulated bilingual environment: A challenge for statistical learning?. *Language Learning and Development*, *5*(1), 30–49.

[27] Winters, S. J., Levi, S. V., Pisoni, D. B. 2008. Identification and discrimination of bilingual talkers across languages. *J. Acoust. Soc. Am. 123*, 4524–4538.

[28] Wong, L. L. N., Liu, S., Han, N., Huang, V. M., Soli, S. D. 2007. Development of two versions of the Mandarin Hearing In Noise Test (MHINT). *Ear & Hearing, 28*(2 Suppl.), 70–74.

[1] Presented as Experiment 2 in [24].

[2] Presented as Experiment 1 in [24].

[3] Only the results from the orthogonal and control blocks will be discussed in this paper.