The acoustic properties of bilingual infant-directed speech

D. Kyle Danielson[a)]
Department of Psychology, The University of British Columbia, 2136 West Mall, Vancouver, British Columbia V6T 1Z4, Canada
kdanielson@psych.ubc.ca

Amanda Seidl
Department of Speech, Language, and Hearing Sciences, Purdue University, Heavilon Hall, West Lafayette, Indiana 47907
aseidl@purdue.edu

Kristine H. Onishi and Golnoush Alamian[b)]
Department of Psychology, McGill University, 1205 avenue du Docteur-Penfield, Montreal, Quebec H3A 1B1, Canada
kris.onishi@mcgill.ca, g.alamian@alumni.ubc.ca

Alejandrina Cristia
Laboratoire de Sciences Cognitives et Psycholinguistique, Centre National de la Recherche Scientifique, Pavillon Jardin, 29 rue d'Ulm, 75005 Paris, France
alecristia@gmail.com

Abstract: Does the acoustic input for bilingual infants equal the conjunction of the input heard by monolinguals of each separate language? The present letter tackles this question, focusing on maternal speech addressed to 11-month-old infants, on the cusp of perceptual attunement. The acoustic characteristics of the point vowels /a,i,u/ were measured in the spontaneous infant-directed speech of French-English bilingual mothers, as well as in the speech of French and English monolingual mothers. Bilingual caregivers produced their two languages with acoustic prosodic separation equal to that of the monolinguals, while also conveying distinct spectral characteristics of the point vowels in their two languages.

PACS numbers: 43.60.Vx, 43.71.Ft, 43.71.Gv, 43.70.Fq [AC]
Date Received: September 30, 2013 **Date Accepted:** December 11, 2013

1. Introduction

Bilingual infants must accomplish the astounding feat of learning two languages in the time it takes monolingual infants to learn one. There is increasing interest in describing the earliest stages of language acquisition, including the extent to which perceptual attunement to the sound categories of the two ambient languages in a bilingual context resembles the attunement found in a monolingual context (Werker, 2012). A key assumption in many descriptions is that bilinguals' input is equal to a conjunction of the two languages. However, this assumption may not always be supported, particularly if a single parent speaks both languages to the child. Indeed, some previous research suggests that the acoustic-phonetics of bilingual speech can be a compromise between the two targets, showing intermediate realizations in certain conditions (Flege *et al.*, 1999;

[a)]Author to whom correspondence should be addressed.
[b)]Current address: Department of Psychiatry, The University of British Columbia, 2255 Wesbrook Mall, Vancouver, BC, V6T 1Z3, Canada.

Pallier *et al.*, 1997). In this letter, we ask if this is also the case of the infant-directed speech produced by balanced bilingual mothers.

We sought to answer two key questions. First, is the distance between the prosodic characteristics of two languages maintained in the infant-directed speech (IDS) produced by bilingual mothers? A positive answer is important, because not only do infants pay more attention to infant- than to adult-directed speech (Soderstrom, 2007), previous infant perception research demonstrates that bilingual infants—perhaps to a greater extent than monolinguals—use prosodic characteristics to discriminate languages from a very early age (Bosch and Sebastiáan-Gallés, 2001; Werker, 2012). Therefore, clear acoustic-prosodic cues could serve as a reference for language separation. Second, is the spectral separation of target vowels for each language maintained in the vowels produced by bilingual caregivers? Indeed, early language acquisition is commonly conceived in terms of adapting to the sound targets of the native language, and IDS has been demonstrated to support the learning of phonetic categories (Werker *et al.*, 2007). In a bilingual context, then, one might expect that the phonetic realizations of vowels in the two languages should resemble the range of those found in monolingual input, so that the bilingual infant may use such a contrast to form the phonemic categories in her two languages. On the other hand, previous work has suggested that some bilingual mothers' vowel formant values are more variable and may lead to more categorical perception errors than those produced by monolingual speakers. Specifically, Bosch and Ramon-Casas (2011) analyzed the Catalan vowel production of bilingual Catalan-Spanish mothers and observed greater variability among those speakers that had been raised in Spanish-speaking homes than those that had been raised in Catalan-speaking homes. However, while that study explored the production of one Catalan vowel contrast in a carefully read speech register, the current study explores the degree to which balanced bilingual caregivers separate their two languages along prosodic dimensions, as well as in a three-way vowel contrast, when speaking in an infant-directed register.

To address these questions, we collected IDS recordings from balanced English-French bilingual caregivers residing in Montreal. We inspected two prosodic characteristics (f_0 and duration), as well as spectral characteristics of point vowels (F1 and F2 in /a/, /i/, and /u/). Significant differences between the languages along these dimensions would indicate that infants in these bilingual homes are receiving potentially separable input from the same caregiver. We also examined the speech of two groups of monolingual speakers: one French and one English. French monolingual speakers were recruited from Montreal and presumably spoke a similar dialect of Quebec French to the bilingual speakers. However, given the great difficulty of recruiting monolingual English-speaking caregivers of 11-month-olds from Montreal with no influence from French,[1] English speakers were instead recruited from the U.S. (Indiana). As the spectral differences between central Indiana and Quebec English instantiations of the vowels of interest are largely confined to differences in vowel backness for /u/ and /a/, we expected little dialect-driven variation between our English samples (Labov *et al.*, 2006). Observations made from these populations allowed us to assess the main underlying assumption questioned above: Whether prosodic and spectral differences between the two languages spoken by a bilingual caregiver are akin to those found when comparing two monolingual caregivers.

2. Methods

2.1 Participants

IDS recordings were obtained from three groups of caregivers with 11-month-olds. Caregivers were native speakers of Quebec French, North American English, or bilingual speakers of both languages. There were 20 mothers in the French group, 22 mothers in the English group, and 9 mothers in the bilingual group. Although the rate of bilingualism is high in Montreal (Statistics Canada, 2011), we applied stringent requirements in order to maintain a sample that exhibited balanced facility in both languages.

Specifically, within the bilingual group, all caregivers reported that their infant was exposed to French and English at least 30% of the time. The caregivers in this group were fluent French-English bilinguals (self-rated fluency out of 9; $M_{French} = 8.4$; $range = 7–9$; $M_{English} = 8.8$; $range = 7–9$) who began learning both languages before the age of 15 (age of first exposure: $M_{French} = 1.5$ yr; $range = 0–10$ yr; $M_{English} = 3$ yr old; $range = 0–14$ yr) and had no exposure to other languages before the age of 15. At the time of testing, mothers used both languages at least 25% of the time and had minimal use of other languages. Our sample was further narrowed, as mothers' speech facility in both languages was confirmed by five naive speakers of each language, who rated the level of foreign accentedness in samples of the mothers' speech. Each of the nine mothers included in our sample scored less than 5.5 (1 = no accent; 10 = heavy accent) in both languages ($M_{\text{French}} = 2.68$; $M_{\text{English}} = 3.63$, $range = 1.56–5.33$).

2.2 Stimuli and procedure

Stimuli consisted of language-specific objects or images designed to elicit words containing the target vowels. For the present study, we focus on the target vowels /a/, /i/, and /u/, preceded by /ʃ/, which were represented in Quebec French with the words *chatte* "female cat," *chiffre* "number," and *choucroute* "sauerkraut"; and in English with the words "shop," "sheep," and "shoes." Caregivers were provided with a clear plastic bag containing three objects/images, two of which were from the same category (e.g., two cats) and a third that was not (e.g., a dog). They were asked to compare and contrast the items in the bag, a task designed to elicit the target items in spontaneous speech. In total, caregivers were asked to describe 30 objects to their children. Here we present the data collected from the point vowels only. The procedure for the monolingual dyads lasted 30–45 min. For the bilingual caregivers, the procedure was divided into two language-specific phases of 30–45 min each, separated by a short break. The order of target languages was counterbalanced across bilingual participants. In this case, to keep the languages as separate as possible, all instructional interactions were conducted in the target language by a native speaker.

Coding and analyses were completed in Praat (Boersma and Weenink, 2012). The tokens in each language were annotated by three highly trained coders who marked the target vowel onset and offset. Following Cristia and Seidl (2013), the onset of the vowel was defined as the first upward crossing after the onset of periodicity following the fricative release. The offset of the vowels was determined as an abrupt attenuation of energy, evident in the waveform and the spectrogram.

Vowels were excluded from analysis if they included background noise, talker overlap, or if the word was whispered or glottalized. Short vowels (<30 ms) and long vowels (>310 ms, two SD above the original mean duration) were excluded from analysis.

Using Praat, fundamental frequency (f_0 in ERB) was measured at the midpoint of each vowel. Formant measurements were estimated using a version of the ceiling optimization algorithm proposed by Escudero *et al.* (2009), at 40% of the vowel's duration (Evanini, 2009), a choice that also limited effects on vowel quality that might have been produced by the varied segments following the target vowels. Because of the nature of the ceiling optimization algorithm, caregivers were excluded from specific target vowel analyses if they did not produce at least two tokens of that vowel.

3. Results[2]

As the data were comprised of spontaneous speech samples, and due to the requirements of the ceiling optimization algorithm, the number of vowel tokens elicited varied across groups. Table 1 gives the number of vowels elicited per language and the number of participants that provided at least two vowel tokens of each vowel.

To determine whether languages differed in the measurements of interest, linear mixed effects regression (LMER) models were used (Baayen, 2008; Bates *et al.*, 2012). Statistical significance is reported here using Markov-chain Monte Carlo

Table 1. Number of vowel tokens elicited per language (number of participants included per vowel).

	/a/	/i/	/u/
Monolingual English	49 (15)	100 (19)	80 (18)
Monolingual French	54 (17)	100 (18)	30 (11)
Bilingual English	31 (9)	48 (9)	33 (8)
Bilingual French	42 (9)	48 (9)	27 (9)

estimated p values (hereafter p) (Baayen, 2011). LMER models were used in order to include individual caregivers in the random effects structure, thus controlling for language-general individual variation orthogonal to the variables of interest.

3.1 Fundamental frequency

In the first model, we declared f_0 at the midpoint of the vowel's duration as the response variable (Table 2), with *language* (levels, as appropriate: monolingual English, monolingual French, bilingual English, bilingual French), *vowel identity* (levels: /a/, /i/, and /u/), and *duration* as factors, and random intercepts declared for individual participants. Including vowel identity and duration in the model allowed for direct observation of the effect of language on fundamental frequency while controlling for these extraneous factors. To serve the current research objective, we report only significant effects of and interactions with language.

When the model's reference level was adjusted to examine the differences between the monolingual speakers, our analysis revealed that French speakers had an overall higher f_0 than English speakers, with duration, vowel identity, and individual variation controlled ($\beta = 0.85$, $SEM = 0.29$, $p = 0.003$). Bilingual mothers similarly separated their two languages on f_0, producing French with a higher f_0 than English ($\beta = 0.64$, $SEM = 0.25$, $p = 0.008$).

3.2 Duration

Linear mixed effects regression models were also fitted to predict vowel duration (Table 2). *Language* (levels, as appropriate: monolingual English, monolingual French, bilingual English, bilingual French), *vowel identity* (levels: /a/, /i/, and /u/), and f_0 (measured at the midpoint of the vowel) were included as factors. Random intercepts were included for individual participants.

Monolingual French vowels were significantly shorter than monolingual English vowels ($\beta = 42.31$, $SEM = 6.70$, $p < 0.001$). Bilingual mothers also separated their two languages by duration, such that their French vowels were shorter than English vowels ($\beta = 64.24$, $SEM = 7.68$, $p < 0.001$).

3.3 First formant

In order to examine the effect of language on the first formant (F1) values of the individual vowels (Fig. 1), separate mixed effects models were fitted to vowel-specific subsets of the data (one model for each vowel /a/, /i/, /u/). In each model, the F1 value (in Hz) was included as the response variable. *Language*, *duration*, and f_0 were included as factors, and random intercepts were computed for individual participants.

Table 2. Mean fundamental frequency and duration by language group across vowels (standard errors in parentheses).

	Monolingual English	Monolingual French	Bilingual English	Bilingual French
f_0 (ERB)	6.52 (0.13)	7.59 (0.33)	6.60 (0.52)	7.06 (0.27)
Duration (ms)	122.54 (6.83)	77.75 (4.71)	127.25 (9.32)	68.99 (2.73)

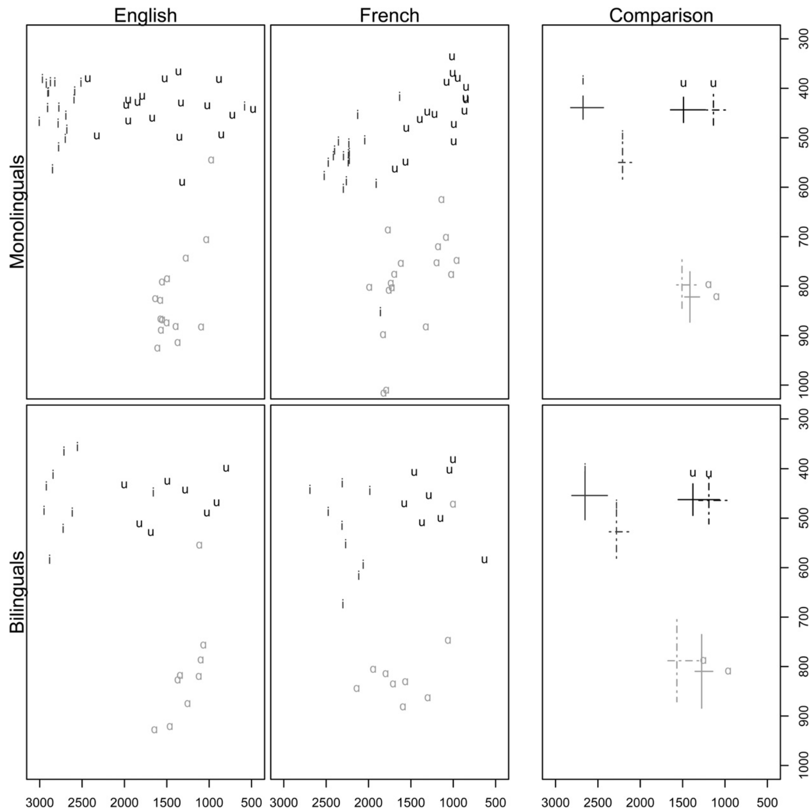

Fig. 1. F1 and F2 values (in Hz) for /i/ (dark gray), /a/ (light gray), and /u/ (black), averaged within speaker (and language) in the monolingual groups (top row) and in the bilingual groups (bottom row). The right-hand panels facilitate the comparison of targets across languages in the two populations. Crosses indicate mean ±2 SEM along each dimension (dotted=French).

The language analysis revealed that there were no significant differences between the F1 values of French and English monolingual speakers when producing /a/ or /u/. Likewise, there were no significant differences in F1 values of English /a/ or /u/ versus French /a/ or /u/ in the bilingual productions.

For /i/, the pattern differs. English monolinguals produced F1 values for /i/ that were significantly lower than those of the French monolinguals ($\beta = 107.03$, $SEM = 20.68$, $p < 0.001$). Likewise, bilinguals produced lower F1 values for /i/ when speaking English than they did when speaking French ($\beta = 97.68$, $SEM = 15.99$, $p < 0.001$).

3.4 Second formant

As was done for F1, separate linear mixed effects models were fitted to the data for each vowel in order to determine the effect of language on the second formant (F2) (Fig. 1). In each model, the F2 value (in Hz) was included as the response variable. *Language*, *vowel duration*, and f_0 were included as factors, and random intercepts were computed for individual participants.

For /i/, English monolinguals produced F2 values that were higher than those of the French monolinguals ($\beta = 440.17$, $SEM = 126.04$, $p < 0.001$). Bilinguals also made a similar distinction, producing higher F2 values for English than for French ($\beta = 406.39$; $SEM = 42.50$, $p < 0.001$).

For /a/, monolingual French and English values did not differ. In contrast, the bilinguals did separate their two languages on this dimension, producing higher F2 values when speaking French than when speaking English ($\beta = 372.89$, $SEM = 47.56$, $p < 0.001$). For /u/, the results were different. Monolingual English speakers produced higher values than French speakers ($\beta = 369.67$, $SEM = 152.98$, $p = 0.017$). In contrast, the bilinguals did not show this distinction, with no significant difference emerging between the F2 values for French and English.

4. Discussion

Bilingual caregivers used unique prosodic profiles in each of their languages, since they produced distinct f_0 and duration patterns in French and English. Moreover, the size of the contrast was similar to that of their monolingual counterparts. This pattern of results is novel. Not only is there a scarcity of evidence on the prosodic realization of IDS as produced by bilingual speakers, but one may also have expected that non-phonemic dimensions, which vary greatly at the individual level, would have been particularly susceptible to merging in the productions of bilingual caregivers. These results demonstrate that distinct prosodic information is available to infants exposed to two languages even through the speech of a single balanced bilingual caregiver, which may enable them to separate their two inputs.

Results from the spectral characteristics of point vowels largely coincide with this picture. We set aside cases in which no difference between languages was found in either the monolingual or the bilingual group, as null effects are difficult to interpret. Cross-linguistic differences were documented for /i/, which was higher and more front in English than in French. There may be many reasons why /i/ was implemented in such a different manner when spoken in the French word *chiffre* versus the English word *sheep*, including item-specific effects, the application of Quebec French closed syllable laxing, and a more general divergence in /i/ implementation across the two languages. As in previous work, the current research suffers from confounding the phonological target with the item, as only one lexical item represented each vowel, a limitation that could be improved upon in future work. Regardless of the source of this divergence, it is important to note that this difference in /i/ was as evident in the bilinguals' two languages as it was in the monolingual reference contrast.

In the second formant measurements of /u/ and /a/, a slightly different pattern emerges. For /u/, monolingual English and monolingual French participants produced significantly different values, while bilingual talkers did not significantly separate their /u/ productions on this dimension. That the monolinguals and bilinguals differed in how separate their /u/ productions were in English and in French may reflect dialect differences rather than a merge of /u/ across languages for the bilinguals, as Quebec and Indiana English likely differ in /u/ backness (Coppler and Pisoni, 2004; Labov *et al.*, 2006; Boberg, 2008).

For /a/, the opposite pattern emerges, with bilinguals exhibiting contrastive F2 patterns that were *not* exhibited by the two groups of monolinguals. Although monolinguals from our two groups do not make this separation, it is possible that bilinguals do so in order to enhance the distinction between the /a/ instantiations in their two languages. Nonetheless, the goal of the present research was not to document specific vowel implementations in these speakers. Our primary aim was to assess whether infants exposed to two languages from a single speaker hear languages that are as different as infants exposed to two languages from different speakers.

To address this question, we investigated prosodic and segmental characteristics of IDS in a group of balanced bilinguals. Significant differences were found in both prosodic parameters and in three out of six spectral parameters measured in point vowels. A comparison with two groups of monolingual speakers revealed that the difference found between the two languages in the speech of bilinguals was as large as that found when comparing two monolingual groups, who also differed along five of the eight dimensions examined. Future work with different designs and populations can elaborate and refine these results. For example, a comparison could be made between the input of bilingual infants hearing two languages from the same speaker (e.g., a bilingual mother) and the input of those hearing speech from two different speakers (e.g., a monolingual mother and a monolingual father speaking different languages), and through observing caregiver-infant interactions in a more naturalistic setting. In the meantime, the present data lend support to the idea that infants exposed to two languages from a single bilingual speaker may be hearing input comparable to the conjunction of input from two monolingual speakers. Therefore, these data serve not only as clarification to researchers studying the environment of infants learning multiple languages, but also to reassure the caregivers of such infants that the input they receive is not compromised.

Acknowledgments

This work was supported in part by funds from NSF 0843959 to A.S. and by ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL* to A.C. A.S. would like to thank Yuanyuan Wang for her assistance in gathering the recordings for this project.

References and links

[1]At the time of the last full Canadian census, only 4.6% of Quebec residents reported speaking only English (with no ability to carry on a conversation in French) (Statistics Canada, 2011). The likelihood of this 4.6% being female, having an 11-month-old infant, and being willing to come into the lab was therefore quite small.

[2]The dataset and additional analyses can be found on the project website at https://sites.google.com/ site/acrsta/Home/nsf_allophones_corpora.

Baayen, R. H. (**2008**). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R* (Cambridge University Press, Cambridge, UK).

Baayen, R. H. (**2011**). "Data sets and functions with 'Analyzing Linguistic Data: A practical introduction to statistics' " (Version 1.4) [R package], http://cran.r-project.org (Last viewed 20 January 2014).

Bates, D., Maechler, M., and Bolker, B. (**2012**). "Linear mixed-effects models using S4 classes" (Version 0.999999-0) [R package], http://cran.r-project.org (Last viewed 20 January 2014).

Boberg, C. (**2008**). "Regional phonetic differentiation in standard Canadian English," J. English Linguist. **36**, 129–154.

Boersma, P., and Weenink, D. (**2012**). *Praat: Doing Phonetics by Computer* (Version 5.3.35), www.praat.org/ (Last viewed December 8, 2012).

Bosch, L., and Ramon-Casas, M. (**2011**). "Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories?," J. Phonetics **39**, 514–526.

Bosch, L., and Sebastiáan-Gallés, N. (**2001**). "Early language discrimination capacities in infants from bilingual environments," in *Research on Child Language Acquisition*, edited by M. Almgren, A. Barrena, M. J. Ezeizabarrena, I. Idiazabal, and B. MacWhinney (Cascadilla Press, Somerville, MA).

Clopper, C. G., and Pisoni, D. B. (**2004**). "Some acoustic cues for the perceptual categorization of American English regional dialects," J. Phonetics **32**, 111–140.

Cristia, A., and Seidl, A. (**2013**). "The hyperarticulation hypothesis of infant-directed speech," J. Child Language. See http://dx.doi.org/10.1017/S0305000912000669.

Escudero, P., Boersma, P., Schurt Rauber, A., and Bion, R. (**2009**). "A cross-dialect acoustic description of vowels: Brazilian and European Portuguese," J. Acoust. Soc. Am. **126**, 1379–1393.

Evanini, K. (**2009**). "The permeability of dialect boundaries: A case study of the region surrounding Erie, Pennsylvania," Publicly Accessible Penn Dissertations, http://repository.upenn.edu/edissertations/86 (Last viewed 20 January 2014).

Flege, J. E., MacKay, I. R. A., and Meador, D. (**1999**). "Native Italian speakers' production and perception of English vowels," J. Acoust. Soc. Am. **106**, 2973–2987.

Labov, W., Ash, S., and Boberg, C. (**2006**). *The Atlas of North American English: Phonetics, Phonology and Sound Change. A Multimedia Reference Tool* (Mouton de Gruyter, Berlin).

Pallier, C., Bosch, L., and Sebastián-Gallés, N. (**1997**). "A limit on behavioral plasticity in speech perception," Cognition, **64**, B9–B17.

Soderstrom, M. (**2007**). "Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants," Dev. Rev. **27**, 501–532.

Statistics Canada (**2011**). "Population by knowledge of official languages, age groups (total), 2011 counts, for Canada, provinces and territories, and census metropolitan areas and census agglomerations," http://www12.statcan.gc.ca (Last viewed 20 January 2014).

Werker, J. F. (**2012**). "Perceptual foundations of bilingual acquisition in infancy," Ann. N.Y. Acad. Sci. **1251**, 50–61.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., and Amano, S. (**2007**). "Infant-directed speech supports phonetic category learning in English and Japanese," Cognition **103**, 147–162.