J. Child Lang. **41** (2014), 935.
doi:10.1017/S0305000914000105

CORRIGENDUM

The hyperarticulation hypothesis of infant-directed speech* – CORRIGENDUM

ALEJANDRINA CRISTIA AND AMANDA SEIDL

doi:10.1017/S0305000912000669, Published online by Cambridge University Press 13 February 2013.

The original version of this article was published containing multiple errors throughout, caused by a problem with the processing of the data. The complete corrected article follows this notice. The changes made do not affect the nature of the results or the basic conclusions.

REFERENCE

Cristia, A. & Seidl, A. (2013). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*. Published online 13 February 2013, doi:10.1017/S0305000912000669.

J. Child Lang. **41** (2014), 913–934.
doi:10.1017/S0305000912000669

The hyperarticulation hypothesis of infant-directed speech*

ALEJANDRINA CRISTIA
Max Planck Institute for Psycholinguistics

AND

AMANDA SEIDL
Purdue University

(*Received* 26 *March* 2012 – *Revised* 23 *July* 2012 – *Accepted* 5 *November* 2012 – *First published online 13 February 2013*)

ABSTRACT

Typically, the point vowels [i,ɑ,u] are acoustically more peripheral in infant-directed speech (IDS) compared to adult-directed speech (ADS). If caregivers seek to highlight lexically relevant contrasts in IDS, then two sounds that are contrastive should become more distinct, whereas two sounds that are surface realizations of the same underlying sound category should not. To test this prediction, vowels that are phonemically contrastive ([i–ɪ] and [eɪ–ɛ]), vowels that map onto the same underlying category ([æ–æ̃] and [ɛ–ɛ̃]), and the point vowels [i,ɑ,u] were elicited in IDS and ADS by American English mothers of two age groups of infants (four- and eleven-month-olds). As in other work, point vowels were produced in more peripheral positions in IDS compared to ADS. However, there was little evidence of

[*] Our deepest appreciation to the families who participated and to Krista Ashmore, Allison Gladfelter, and Kat Montoya for the phonetic coding. Portions of this work were discussed at the 2010 Journees des Etudes sur la Parole, the Acoustical Society of America 2010 November meeting, and the 2011 Testing Models of Phonetics and Phonology Workshop at the University of Colorado at Boulder. We are grateful to these audiences and our colleagues at Purdue University, the Laboratoire de Sciences Cognitives et Psycholinguistique, the University of Pennsylvania, and Northwestern University for helpful discussions; as well as to Titia Benders, two anonymous reviewers, and the editors for insightful comments and careful reading of a previous version of this manuscript. We also thank Kristine H. Onishi for help in the design of the elicitation material. All errors are our own. AS collected the recordings, and supervised the coders; AC generated the coder training materials and performed the acoustical and statistical analyses. Both authors contributed to the writing. This work was supported by funds from NSF 0843959 to Amanda Seidl. AC gratefully acknowledges the financial support of the Fondation Fyssen and Ecole de Neurosciences de Paris. Address for correspondence: Alejandrina Cristia, Neurobiology of Language, Max Planck Institute for Psycholinguistics, PO Box 310, 6500AH, Nijmegen, Netherlands. e-mail: alecristia@gmail.com

hyperarticulation per se (e.g. [i–ɪ] was *hypo*articulated). We suggest that across-the-board lexically based hyperarticulation is not a necessary feature of IDS.

INTRODUCTION

The speech we hear in infancy has an indelible effect on our perception (Dupoux, Peperkamp & Sebastián-Gallés, 2010). In many cultures, much of this early input comes from a few primary caregivers, often the mother. Although a child hears her mother speaking to others (van de Weijer, 1998), experimental evidence demonstrates that speech directly addressed to infants has an inordinate influence on the child (see a recent summary in Soderstrom, 2007), such that the segmental and suprasegmental characteristics of maternal infant-directed speech could be key determinants of lifelong patterns of perception. Influential work suggests that caregivers produce vowels (Bernstein Ratner, 1984; Burnham, Kitamura & Vollmer-Conna, 2003; Kuhl *et al.*, 1997; Uther, Knoll & Burnham, 2007) and tones (Liu, Tsao & Kuhl, 2007, 2009) more clearly when addressing their infant listeners than when addressing other adults. For example, Kuhl *et al.* (1997) document that Russian-, American English-, and Swedish-speaking mothers all produced point vowels [i,ɑ,u] that were more extreme in the F1–F2 acoustic space in infant-directed speech (IDS) than in adult-directed speech (ADS). These findings were extended to Australian English (Burnham *et al.*, 2003) and Mandarin Chinese (Liu, Kuhl & Tsao, 2003). Such evidence has led to the characterization of IDS as *hyperspeech* (Fernald, 2000), and to the general proposal that talkers have an implicit goal of promoting language acquisition by highlighting phonemically contrastive categories (Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola & Nelson, 2008; Liu *et al.*, 2003). For example, in their seminal paper, Kuhl *et al.* (1997: 686) argue that '[t]he exaggerated form serves two functions: It more effectively separates sounds into contrasting categories, and it highlights the parameters on which speech categories are distinguished and by which speech can be imitated by the child'. We will call this position the *hyperarticulation* hypothesis.

Although vowel space expansion is often highlighted in literature reviews on IDS (Cristia, in press), not all acoustic-phonetic studies report enhancement of sound categories. For example, a study with Norwegian mothers (Englund & Behne, 2006) did not find expanded point vowels in speech to infants. Moreover, investigation of specific phonemic contrasts, such as vowel quality and vowel length (Englund & Behne, 2005) and a variety of consonantal contrasts (see a recent discussion in Cristia, 2010), has thus far yielded mixed results. One apparent source of variance is the child's age (or linguistic development; Englund & Behne, 2006; Kitamura & Burnham, 2003; Kitamura, Thanavishuth, Burnham & Luksaneeyanawin, 2002; Malsheen, 1980).

In terms of segmental realization, enhancement is more frequently reported in speech to older infants than in speech to younger infants (e.g. Bernstein Ratner, 1984; Cristia, 2010). Such findings are not incompatible with the hyperarticulation hypothesis, insofar as caregivers are more inclined to hyperarticulate when the child is actively learning words.

Nonetheless, the variability in findings casts an empirical shadow over the hyperarticulation hypothesis. From a theoretical viewpoint, it should be conceded that the talker's implicit intention to hyperarticulate is only one of several possible factors that would result in vowel space expansion, which could in fact be a side effect of one or more of the many differences between the registers (as is often used in IDS literature, we use the word 'register' as a shorthand to refer to IDS and ADS; e.g. Englund & Behne, 2005; Liu *et al.*, 2007). For example, IDS consists of shorter phrases (Fernald, Taeschner, Dunn, Papousek, Boysson-Bardies & Fukui, 1989), in which proportionately more sounds should be strengthened by virtue of their alignment with strong prosodic boundaries (Gendrot & Gerdes, 2009; Keating, Cho, Fougeron & Hsu, 2003). Additionally, preboundary syllables are longer in IDS than ADS (e.g. Lam & Kitamura, 2010), and speech rate may be slower in IDS (Englund & Behne, 2005; Fernald *et al.*, 1989), which may make all vowels less likely to be undershot (Gendrot & Adda-Decker, 2007; Lindblom, 1963). This hypothesis seems plausible given that phonetic research on ADS suggests that a vowel is more likely to be produced in more peripheral positions when it is longer (Gendrot & Adda-Decker, 2007; Lindblom, 1963) and closer to a large prosodic boundary (Cho, 2005; Fougeron, 2001). Thus, since there are more large prosodic boundaries and vowels are longer in IDS, it is possible that at least some of the spectral enhancement found for point vowels is a side effect of other articulatory behaviors. There are several additional characteristics of IDS that could affect phonetic implementation. For example, smiling has complex acoustic consequences (Tartter, 1980), and a recent study documents that smiling increases F2 for [i] and [ɑ] but not [u] (Fagel, 2010). If caregivers smile more to their infants than to adults, this difference in smiling could result in an expansion of vowel space purely due to the differential effects of smiling on the different vowels. In general, one could postulate a SIDE-EFFECTS EXPLANATION according to which hyperarticulation is not an implicit goal of caregivers, but vowel space expansion is a potential outcome of other behaviors, such as smiling and producing shorter and slower utterances.

In view of the existence of conflicting evidence, and given that other potential explanations exist, it seems premature to conclude that vowel space expansion occurs due to hyperarticulation, and that hyperarticulation is a key feature of IDS. However, the side-effects explanation is rather under-specified, lacking the predictive power of the hyperarticulation hypothesis.

Therefore, the present study was designed to seek evidence which would directly speak to the hyperarticulation hypothesis, by testing a prediction that is unique to this view. It has been stated that hyperarticulation serves two goals: to separate contrasting categories, and to highlight dimensions that should be encoded by the child when imitating (Kuhl *et al.*, 1997: 686). The strongest version of this lexically driven hyperarticulation hypothesis uniquely predicts that sounds that are lexically contrastive (PHONEMIC sets) will be hyperarticulated in IDS compared to ADS, but sounds that are contextually determined (ALLOPHONIC sets) will not be hyperarticulated.

All languages have allophones, two or more surface instantiations of the same underlying sound category, and both phonemic and allophonic sets are present in infants' input. For example, English-learning infants encounter tense and lax, as well as nasal and oral, vowels in their input. However, vowel tenseness and vowel nasality differ in their phonological status. Tenseness is phonemic in English because minimal pairs can exist along the tenseness dimension; e.g. *beet–bit* /bit–bɪt/. In contrast, vowel nasality is an allophonic dimension, since nasal vowels are positional variants of oral vowels, typically surfacing before nasal consonants; for example, the vowels in *man* and *bat* differ in nasality, but they map onto the same underlying category in English. Since only phonemic dimensions are contrastive, and need to be explicitly encoded when imitating, all phonemic (but no allophonic) dimensions should be targets of hyperarticulation in IDS. It may be noted that the underlying mechanism that would mediate such selective hyperarticulation need not be conscious, as speakers implicitly enhance the acoustic dimensions that are relevant for the phonological distinction being made (Kang & Guion, 2008).

We measured phonemic and allophonic sets in English-speaking mothers; in particular, we elicited the vowel tenseness sets [eɪ–ɛ] and [i–ɪ] (phonemic in English), and the vowel nasality sets [æ–æ̃] and [ɛ–ɛ̃] (allophonic in English). Two different vowel qualities were used to represent each set to seek an internal replication; if enhancement follows from phonological status, it should happen in both qualities representing tenseness, and in neither quality representing nasalization. In view of arguments that enhancement is partially dependent on the child's age, we recorded caregivers talking to two different age groups of infants, one prelexical and the other beginning to acquire a lexicon.

Based on the hyperarticulation hypothesis, one can expect three outcomes. First, vowel spaces should be larger in IDS than in ADS, particularly in speech to older infants. Second, the distance between sounds differing in tenseness (a phonemic dimension) should be greater in IDS than ADS, particularly in speech to older infants. Third, the distance between vowels differing in nasality (an allophonic dimension here) should not be expanded between IDS and ADS. To convey such differences across registers in a way that clearly encodes these predictions, we will refer to cases in which distances are larger in IDS than in ADS as ENHANCEMENT; if distances are smaller in IDS than ADS, we will call this DETERIORATION. We would like to stress that the present study was designed to assess lexically driven hyperarticulation. The additional hypotheses for IDS expansion outlined above (e.g. the side-effects hypothesis) are also worthy of study, but they lie beyond the scope of the present study. Therefore, we will limit our discussion to whether predictions made from the lexically driven hyperarticulation hypothesis are supported or not. Specifically, an expansion of the distance between vowels differing along a phonemic dimension, but not between vowels differing along an allophonic dimension, would constitute clearer evidence in favor of the hyperarticulation hypothesis because this view neatly predicts that pattern of results.

METHODS

Participants

We tagged and analyzed the speech of eighteen mothers of eleven-month-olds (mean = 0;11·3, range = 0;11·0 to 0;12·0; girls = 7, boys = 11), and twenty-eight mothers of four-month-olds (mean = 0;4·15, range = 0;3·29 to 0;4·27; girls = 16, boys = 12). Henceforth, we will refer to the talkers addressing eleven-month-olds as the '11M' group; and those speaking to four-month-olds, the '4M' group. More infants were recruited for the 4M group because null effects in previous literature investigating this age group would indicate that more power is desirable. All of the mothers were native American English speakers and they were highly educated. Half of them had a bachelor's degree or greater (11M: mean = 16·2, *SD* = 2·6, range = 12 to 22; 4M: mean = 15·9, *SD* = 2·7, range = 10 to 22). Only three previous studies on hyperarticulation have reported levels of education of the caregivers recorded, all of them on Taiwanese caregivers (Liu *et al.*, 2003; 2007; 2009); as in our sample, those caregivers were highly educated. For such highly educated samples, variation in education does not relate to hyperarticulation. Liu *et al.* (2003) found no correlation between caregiver's speech and education, income, or occupation. Thirteen additional caregivers could not be included for the following reasons: data loss or poor quality of the recording (5); caregiver was male (4); adult-directed speech could not be collected due to child fussiness (2); either the parent or the child had a disorder (2). Families received a book or toy as well as a diploma for their participation. All caregivers signed an informed consent form, and this research was approved by the university IRB.

TABLE 1. *Lexical items used in the elicitation stimuli*

Phonemic							
[eɪ]	[ɛ]	[eɪ]	[ɛ]	[eɪ]	[ɛ]	[eɪ]	[ɛ]
bacon	pegboard	basil	pesto	daycare	decker	paper	pepsi[1]
[i]	[ɪ]	[i]	[ɪ]	[i]	[ɪ]	[i]	[ɪ]
beetle	piglet[2]	peekaboo	picnic	teakettle	dictionary	teaspoon	disney
Allophonic							
[ɛ]	[ɛ̃]	[ɛ]	[ɛ̃]	[ɛ]	[ɛ̃]	[ɛ]	[ɛ̃]
pepsi[1]	Benji	best	pencil	pedal	pendant	Teddy	tender
[æ]	[æ̃]	[æ]	[æ̃]	[æ]	[æ̃]	[æ]	[æ̃]
baboon	bamboo	tassle	dancer	bassinet	pansy	tapioca	tambourine

NOTES: [1] *pepsi* served as both an oral vowel, and a lax vowel. [2] This item was *pistachio* for the first nineteen participants; it was thereafter changed to *piglet* because *pistachio* was variably realized. In the first nineteen participants, only correct /ɪ/ instances were coded.

Recordings

Recordings took place in a small, sound-attenuated room. The caregiver was fitted with a Lavalier microphone (AKG WMS40 Pro Presenter Set Flexx UHF Diversity CK55), whose signal was recorded onto a Marantz Professional Solid State Recorder (PMD660ENG). The caregiver was told that we were interested in how caregivers talk about categories to their infants. She was given a box full of smaller bags, each containing object/picture sets that had some reference to a target word, which was written out on the bag. The complete list of target words for tenseness and nasality is shown in Table 1; in addition, one item was included to represent each point vowel [i,ɑ,u]: *sheep*, *shop*, *shoes*. For each target word, there were two similar objects/pictures (e.g. two picnic baskets of different sizes and colors) and one oddball object (a picnic blanket). Thus, the caregiver most often told her infant that the two similar objects went together and the other was an oddball. The caregiver was left alone with the child for about 30–45 minutes, until she had described all the object sets to their child. Then, an experimenter and a confederate entered the room, and the caregiver was asked to perform the same task to the new adult confederate, who (caregivers were told) had never seen the objects before.

Stimuli

The tenseness contrasts were both front because pilot analyses revealed more measurement errors for back vowels; furthermore, including a contrast where one of the categories was a point vowel increased the comparability of the tenseness contrasts with the vowel triangle. The specific nasality contrasts [æ–æ̃] and [ɛ–ɛ̃] were selected as the closest to extant French nasal–oral vowel contrasts, foreseeing a cross-linguistic comparison. Previous work typically uses a single lexical item to represent each sound under study. In order to reduce the impact that this choice could have on our results, we preferred to have each contrast represented by four pairs of lexical items. We favored imageable objects; and pairs were matched as closely as possible in a number of phonological variables (e.g. preceding and following context, word length) and lexical factors (lexical frequency). We return to this in the 'Discussion'.

Coding and measurement

Coding and acoustic analyses were conducted in Praat (Boersma & Weenink, 2005). Native speakers of English listened to the files and marked the approximate location of the syllable containing the target vowel. Coder training materials, scripts, color versions of figures, and further analyses are provided on the project website (Cristia, 2011). Three highly trained coders then marked the target vowel onset and offset for all of the syllables; at a second and third stage, all of the coding was checked by a single coder, to ensure that there was perfect consonance in criteria. The onset of the vowel was defined as the first upward crossing after the onset of periodicity following the burst or fricative release of the preceding consonant. The offset of the vowels was determined as an abrupt attenuation of energy, evident in both the waveform and the spectrogram. Coders were particularly careful when determining the offset of a vowel followed by a nasal consonant, which is considered difficult when the vowel is nasalized. In these cases, all sources of information (attenuation and increase in regularity in the waveform; sudden appearance of nasal zeros in the spectrogram; and auditory feedback) were used, and all cases of disagreement were decided by consensus. Vowels were not tagged, coded, or analyzed if the formant structure was affected by background noise or talker overlap, or if the word had been whispered or heavily glottalized. Tagged vowels shorter than 40 ms were excluded at the analysis stage, as the window of analysis used for nasality would then include neighboring sounds. This minimum length criterion was extended to the tenseness contrasts and to the point vowels, to keep the vowel samples comparable.

Measures of duration, fundamental frequency, and first and second formants (F1 and F2) were extracted from the point vowels, which served as a control of the quality of IDS. For tenseness, we measured vowel duration (and applied the logarithm, as in e.g. Escudero, Boersma, Rauber & Bion, 2009) and F1 and F2 frequency (e.g. Hillenbrand, Getty, Clark & Wheeler, 1995). F1 and F2 were measured at both 40% and 80% of the vowel duration (using an implementation of the ceiling optimization algorithm proposed by Escudero *et al.*, 2009, which is robust to talker variability; see the project website for more details), since some work suggests that vowel-internal formant changes improve categorization, particularly for

https://doi.org/10.1017/S0305000914000105 Published online by Cambridge University Press

diphthongs like [eɪ] (e.g. Hillenbrand, Clark & Nearey, 2001). Tense vowels are longer and have more peripheral formant frequencies than their lax counterparts (in the vowels studied here, lower F1 and higher F2).

For nasality, we measured F1 bandwidth and the difference in amplitude between the first formant and the first (P0) and second (P1) nasal poles, measures known as A1–P0 and A1–P1 (e.g., Chen, 1997). Our algorithm follows closely that proposed in Pruthi (2007; see the project website for further details and comparisons with other algorithms). The loudest pole below the first formant or immediately above F1 was assigned to P0, and the loudest pole between F1 and F2 was assigned to P1. Since these measurements rely on accurate F1 tracking, F1 frequency was calculated using the same algorithm mentioned above prior to pole tracking. As with tenseness, these measures were gathered at 40% and 80% of the vowel duration. Nasal vowels have larger F1 bandwidths and lower A1–P0 and A1–P1 than their oral counterparts (due to both an increase in the intensity of the nasal poles, and a decrease in the intensity of the first formant). The choice made came from well-established automatic algorithms; gathering production data or perception data was impractical given the size of this corpus.

Analyses and distance calculations

We assessed IDS quality in terms of vowel duration, fundamental frequency height, and vowel space size using the point vowel tokens. Four caregivers did not provide any usable point vowel tokens in one of the registers, and were thus excluded from the analyses of IDS quality. Vowel space was calculated as the area within the triangle whose vertices correspond to the average F1 and F2 values for each vowel [i,ɑ,u] (as in Kuhl *et al.*, 1997).

With respect to the distance spanned along phonemic and allophonic dimensions, acoustic-phonetic measurements were combined into two estimates of contrastiveness. The first measure was simply the Euclidean distance between the centroids of the categories in the multidimensional space defined by the relevant acoustic measurements. This estimate, akin to the separation between the centroid of, e.g., [eɪ] and [ɛ], is highly comparable to measures of separation used in previous IDS research (Englund & Behne, 2005; Kondaurova, Bergeson & Dilley, 2012). However, perceptual category learning research suggests that contrastiveness is a function not only of the separation between two categories, but also the variability within the individual categories (e.g. Markman & Maddox, 2003). Imagine two cases, A and B. In case A, the centers of two sound categories are separated by 10 units in some acoustic dimension, and the standard deviation of the categories is 5 units. Categories in case B are also separated by 10 units, but the standard deviation is 50. In consequence, it

https://doi.org/10.1017/S0305000914000105 Published online by Cambridge University Press

should be much harder to learn the categories in case B than in case A.[1] Existing evidence confirms the relevance of both separation and within-category variability (along identifying dimensions) in adult listeners' perception of speech sounds. For example, all else being equal, adults are slower and more error-prone when classifying /s/ and /ʃ/ spoken by talkers that pronounce them similarly; or when the within-category variability is large (Newman, Clouse & Burnham, 2001). Therefore, a second estimate was calculated to assess how overlapping or well-separated two categories are in the speech of individual talkers, namely D(a) (e.g. Newman *et al.*, 2001). This estimate is the between-category distance (the Euclidean distance above) divided by the within-category variance. To ensure that the measures of distance obtained were representative of the talker and population, we excluded tokens with measurement values that were spurious; and we excluded talkers who had few tokens per category, as explained in more detail below.

In terms of spurious tokens, there were 170 nasal/oral tokens that were outliers in F1 bandwidth (larger than 500 Hz); 92 [i] tokens that had outlier values for F2 (lower than 2000 Hz); 20 [ɪ] tokens with F2 under 1500 Hz; 16 [eɪ] tokens with F2 under 1500 Hz; and 19 [ɛ] tokens under 1250 Hz. In addition, 10 [i] tokens with F2 under 2000 and 2 [u] tokens with F1 over 1500 Hz were excluded from the vowel space calculations. These cut-offs were selected on the basis of visual inspection of vowel plots, available online. After these exclusions, there were still well over 500 tokens for each type. Specifically, the counts by register (ADS–IDS) for tenseness were: [eɪ] 442–500; [ɛ] 296–389; [i] 296–605; [ɪ] 300–445. For nasality, the counts were: [æ] 395–469; [æ̃] 434–505; [ɛ] 318–441; [ɛ̃] 298–414.

As for the criterion to exclude talkers on the basis of having few tokens, we calculated distances setting the minimum tokens of each category in each register at every integral between 2 and 6. The same pattern of results ensued for ALL minimum criteria; boxplots showing the stability of results as a function of the minima can be found on the project website. In the 'Results' section, we focus on the analysis where the minimum token requirement was set at 4 (i.e. to be included for [i–ɪ], a talker had to have 4 [i] in IDS, 4 [i] in ADS, 4 [ɪ] in IDS and 4 [ɪ] in ADS), as a compromise between having enough tokens to calculate the D(a) reliably within individual talkers, and including as many talkers as possible in the analyses of individual vowel pairs. A total of 27 talkers (16 mothers of 4M and 11 of 11M) for [eɪ–ɛ]; 29 talkers (18 4M, 11 11M) for [i–ɪ]; 36 talkers

[1] These considerations relate to variation along the same dimension that separates the categories. Naturally, variation along irrelevant (and separable) dimensions need not impact discrimination negatively. Indeed, some work suggests that infant and adult learners can even profit from clearly irrelevant variability when learning words or foreign categories (e.g. Lively, Logan & Pisoni, 1993; Rost & McMurray, 2009).

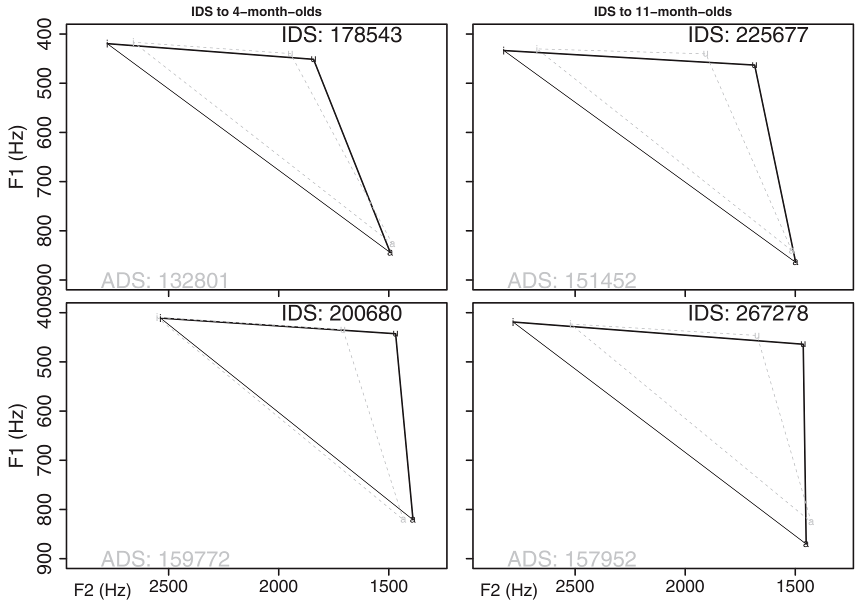

Fig. 1. Vowel space in speech addressed to infants (black, solid) and adults (gray, dashed). The top row shows the averages for the measurement gathered 40% into the vowel, the bottom row those corresponding to the 80% measurement. Graphs on the left side correspond to talkers addressing four-month-olds; on the right, those addressing eleven-month-olds. The numbers in the corners indicate the average vowel space area in Hz^2.

(20 4M, 16 11M) for [æ–æ̃]; and 25 talkers (14 4M, 11 11M) for [ɛ–ɛ̃] were included in our analyses. Notice that these sample sizes are still considerable, as few previous IDS studies have had more than 10 talkers per group. The distribution of the distance measures did not meet the requirements for parametric statistical testing. Therefore, a sign test was used to compare the means for IDS and ADS for the 4 cases (2 vowel qualities × 2 dimensions) for each of the two distance measurements in each age group. The critical α was set to .006 to control for the 8 comparisons.

RESULTS

IDS prosody and vowel space size

Before moving on to the main results, it is important to demonstrate that our recordings have truly captured the two registers. Specifically, the segmental and suprasegmental differences between our IDS and ADS samples should be comparable to the differences reported in previous hyperarticulation work. As shown in Figure 1, vowel spaces were larger in IDS than in ADS;

https://doi.org/10.1017/S0305000914000105 Published online by Cambridge University Press

these spaces were calculated collapsing across all talkers, since few talkers produced sufficient tokens to estimate individual vowel spaces reliably. Since IDS characteristics did not vary across age groups in the current sample (no effect of age group in a repeated measure ANOVA), all of the following averages are provided collapsing across groups, to simplify the exposition. The ratio of the vowel space size in IDS divided by the vowel space size in ADS is about 1·5 in the current sample, which is well within the range of ratios found in previous hyperarticulation research (Kuhl *et al.*, 1997: 1·9; Lam & Kitamura, 2010: 1·2; Liu *et al.*, 2009: 1·3; Uther *et al.*, 2007: 2·2). The ratio of vowel duration was 1·4 for point vowels (163 ms in IDS vs. 120 ms in ADS), and 1·2 in the tense/lax vowels (89 ms in IDS vs. 74 ms in ADS) and the nasal/oral (96 ms in IDS vs. 81 ms in ADS) vowels. These ratios are remarkably similar to those reported in previous hyperarticulation work (Kuhl *et al.*, 1997: not reported; Lam & Kitamura, 2010: 1·4; Liu *et al.*, 2009: 1·5; Uther *et al.*, 2007: 1·4). Average pitch was measured in ERB, a psychoacoustically based scale. Using this scale, the ratio was 1·2 for point vowels (6·8 ERB in IDS vs. 5·7 ERB in ADS), 1·2 in tense/lax vowels (6·5 ERB in IDS vs. 5·5 ERB in ADS), and 1·1 in nasal/oral vowels (6·1 ERB in IDS vs. 5·4 ERB in ADS). These ratios were comparable to one previous study (Liu *et al.*, 2009: 1·3), although other studies report higher ratios (Kuhl *et al.*, 1997: not reported; Lam & Kitamura, 2010: 1·85; Uther *et al.*, 2007: 1·8). We suspect that the somewhat lower ratio we report here is due to all of these previous studies reporting pitch in hertz. When we convert our averages to hertz, we find higher ratios; e.g. 1·3 for point vowels. In brief, our sample exhibits vowel space expansion, longer vowel durations, and higher average pitch, with differences between IDS and ADS that are comparable to the speech samples that, in previous research, have been used to support the hyperarticulation view of IDS.

Phonetic enhancement

The main goal of this article is to compare between-category distances in IDS with those in ADS spoken by the same talker. As mentioned above, enhancement occurs when distances are significantly LARGER in IDS than in ADS, and deterioration when they are significantly SMALLER in IDS than in ADS. Most previous work has operationalized contrastiveness as the separation between category centers, but D(a) (the between-category distances divided by the within-category variance) may yield a psychologically more accurate representation than pure distance. Boxplots of distances are shown in Figure 2, which complement the non-parametric statistical analyses presented in the following paragraph. When inspecting this figure, bear in mind that, if caregivers produced all categories more

https://doi.org/10.1017/S0305000914000105 Published online by Cambridge University Press

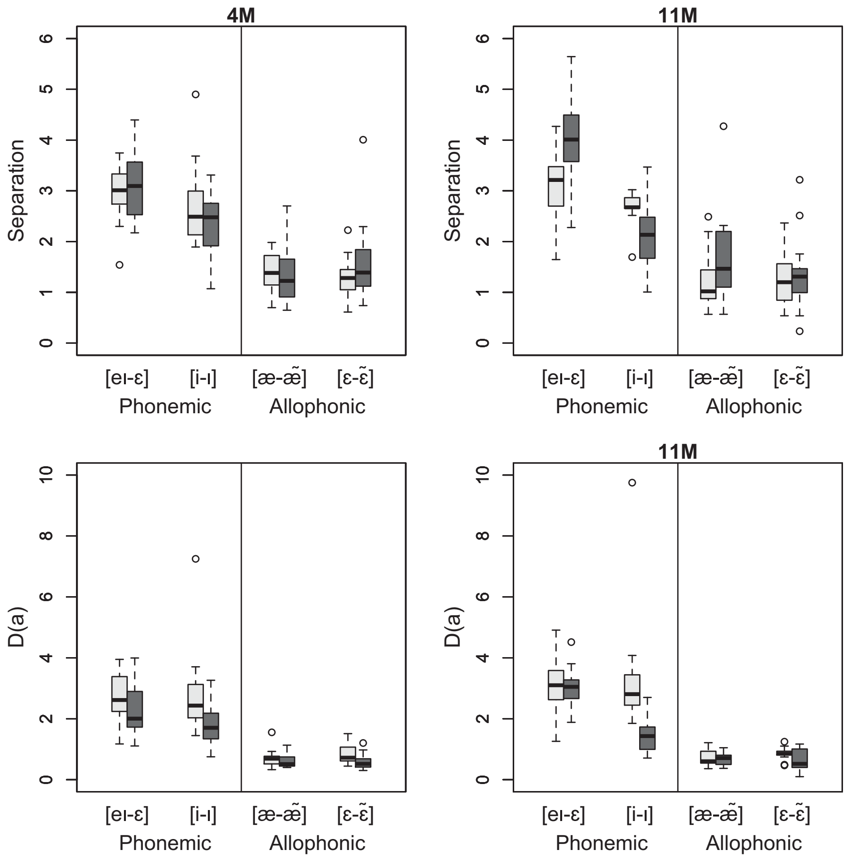

Fig. 2. Boxplots of contrastiveness by vowel set and register (adult-directed speech in light gray, infant-directed speech in dark gray) in terms of the separation between category centers or separation (top panels); and in terms of D(a), the between-categories distance divided by the within-category variance (bottom panels). The panels on the left show data from the 4M group; the right panels, those from the 11M group. Units are akin to standard deviations (see main text for details).

clearly in IDS than ADS, dark boxes should be higher than light ones. Given that previous work has found enhancement to change with age, we discuss results within each age group in order.

Speech to younger infants. When separation was considered, there were no consistent patterns for [eɪ–ɛ], but there was a trend towards deterioration for [i–ɪ] (13 out of 18 mothers had greater distances in ADS than IDS). As for vowels differing along the allophonic dimension of nasalization, there were no consistent patterns of enhancement or deterioration for either vowel quality. When D(a) was considered, there was a trend towards deterioration for [eɪ–ɛ] (11 out of 16 4M showed greater distances in ADS than IDS), which was significant for [i–ɪ] (for 17 out of 18 caregivers the D(a) was larger in ADS than in IDS; $p < \cdot 001$). No consistent pattern was evident in this measure for [ɛ–ɛ̃] or [æ–æ̃].

Speech to older infants. When separation was considered, opposite patterns were found for the two vowel sets differing along the phonemic dimension of tenseness. Specifically, [eɪ–ɛ] was somewhat enhanced, with 9 out of 11 caregivers showing larger distances for IDS as compared with ADS ($p = \cdot 05$), but there was a trend towards deterioration in [i–ɪ] (8 out of 11 caregivers exhibited larger distances for ADS as compared to IDS). None of the tests involving the sets differing along the allophonic dimension approached significance. When D(a) was considered, caregivers were inconsistent in the relative changes of distance across registers for [eɪ–ɛ], but there was significant deterioration for [i–ɪ], since all 11 caregivers of 11M had greater distances in ADS than in IDS ($p < \cdot 001$). No significant results emerged for the allophonic sets (although most caregivers showed smaller D(a) in IDS for [æ–æ̃]; 12 out of 15 caregivers).

Given that a somewhat different pattern of results was found in separation and D(a)s, it was necessary to illustrate the dimension along which the distance measures differ, namely within-category variance. A scatterplot of the four vowels used for the tenseness contrasts is shown in Figure 3. This plot shows that the greater variability in IDS is not simply due to a greater number of outliers, and that it can result in greater overlap between categories in IDS. Interestingly, post hoc analyses revealed that the larger within-category variability in IDS, calculated for each vowel type separately, was highly significant. Since our study was not specifically designed to measure variability (e.g., different number of tokens are available for IDS than ADS), we do not report them here, but these analyses are available in the project website.

DISCUSSION

Previous work has often reported that an enlargement of vowel space in IDS is evidence of hyperarticulation. This assessment may be based on the intuition that caregivers implicitly attempt to promote language acquisition by highlighting contrasts for their infants. However, more extreme point vowels could ensue from factors other than hyperarticulation. To directly examine the hyperarticulation characterization of IDS, we measured vowels differing along phonemic and allophonic dimensions in both registers (IDS, ADS), in addition to vowel space size. As in much previous work, caregivers recorded in the present study produced point vowels [i,ɑ,u] with more peripheral acoustic characteristics. We predicted

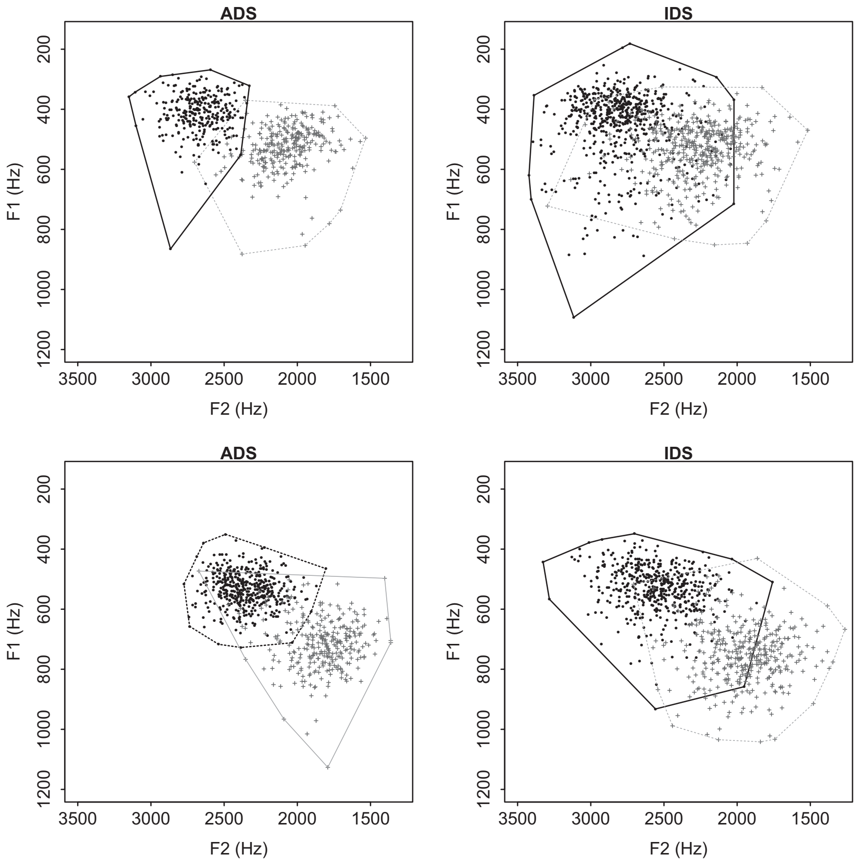

Fig. 3. Scatterplots of [i–ɪ] (top) and [eɪ–ɛ] (bottom) by register (ADS on the left, IDS on the right) along F1 and F2 at 40% of the vowel duration. Tense vowels are shown in black filled circles, with a solid line marking the outline defined by the most extreme values; lax vowels in gray crosses, with a dashed line marking the outline.

that if this was due to an implicit goal to highlight lexically contrastive sounds, then sounds differing along other phonemically contrastive dimensions, such as those differing on tenseness in English, should also show expansion. Based on previous work, we also expected this pattern to be more marked in speech to eleven-month-olds than in speech to younger infants.

Results for the four-month-old group yielded little support for the prediction that phonemic contrasts would be enhanced when simple category separation was considered; moreover, there was evidence to the contrary when within-category variance was also taken into account. Contrary to

https://doi.org/10.1017/S0305000914000105 Published online by Cambridge University Press

our predictions, results for the older group did not provide greater support for the hyperarticulation hypothesis. In speech to eleven-month-olds, [eɪ–ɛ] tended to be produced with larger between-category separation in IDS, but [i–ɪ] trended in the opposite direction. Furthermore, the enhancement found for [eɪ–ɛ] disappeared when within-category variance was included in distance calculations and the deterioration of [i–ɪ] became more pronounced.

Our second set of predictions involved sounds that differed along an allophonic dimension. We reasoned that, if enhancement was aimed at exaggerating only lexically relevant contrasts, then there would be no such enhancement in vowels that were allophones of the same underlying category. In general, no enhancement was evident for these nasality contrasts.

Although at first glance one may view the latter allophonic findings as fitting the predictions of the lexically based hyperarticulation hypothesis, such an interpretation rests on the assumption that enhancement would be found for the phonemic contrasts. Given that this premise was not met, the null results for the nasality contrasts are difficult to interpret. We chose to report them here, nonetheless, because they expand our knowledge of IDS in an unprecedented dimension. We hope that future work may be able to shed light on why some vowel contrasts pattern in one way (there was a clear tendency for deterioration in [i–ɪ]) or another (there was a trend towards greater separation, which was compensated by greater variability, in [eɪ–ɛ]), whereas other contrasts exhibit less consistent changes across registers (there was no clear change for [ɛ–ɛ̃]). These data suffice to test the hypothesis we set out to test, but clearly they do not solve all extant questions on IDS, including which specific factors lead to hyper- or hypoarticulation in this register.

While the present results lend little support to a lexically driven hyperarticulation view of IDS, whereby caregivers (implicitly) promote phonological acquisition through the expansion of lexically relevant phonological categories, they do not negate the existence of vowel space, pitch average, range, and duration expansion. Instead, these results suggest only that the STRONGEST version of the hyperarticulation hypothesis is false: It is simply not the case that ALL phonemic contrasts are enhanced in IDS. Weaker versions of the hyperarticulation hypothesis could be put forward that are more or less general. It could still be the case that caregivers enhance the point vowels to promote language acquisition, but do not enhance any other vowel contrasts; or perhaps caregivers only increase the separation between category centers of front non-high vowels.

Nonetheless, it appears to us that recent results involving the comparison of IDS to infants with a hearing impairment cast additional shadows over the hyperarticulation view of IDS. Lam and Kitamura (2010) studied speech directed to a set of twins, where one had normal hearing and the other did not. Their mother produced a larger vowel space only in speech addressed to the normal hearing child, even though the hearing-impaired child was probably in greater need of hyperarticulation.

A subsequent experimental manipulation used a double-video set-up, where infant and mother interact with each other through a closed video circuit, to further investigate how hearing status affects vowel space expansions (Lam & Kitamura, 2012). In one condition the infant could hear the caregiver perfectly, whereas in other conditions the audio return to the infant was turned down or off. Acoustic analyses revealed that mothers expanded their vowel spaces only in the condition where the infant could hear them perfectly. The study added a second manipulation, which was crossed with hearing condition: the explicit belief that the infant could or could not hear. Thus, sometimes caregivers were told that their infant could hear them when in fact the infant could not, and other times they were correctly informed that their infant could not hear. In fact, telling the mothers that the infant could or could not hear them did not affect vowel space expansion, and did not interact with hearing status. Consequently, it was not the talkers' belief about the situation, but rather something about the infant's response which shaped caregivers' vowel space in real time. In other words, caregivers of both normal-hearing and (temporarily) hearing-impaired infants likely want their infant to acquire language; if caregivers differ in their vowel space expansion, this expansion must follow from some other variable.

The debate between the view that caregivers naturally display behavior that is directed towards promoting language acquisition and those who hold less directed views of caregivers' speech is an old one (see, for example, Snow & Ferguson, 1977). Together with other research, the present study contributes to the push towards replacing the hyperarticulation view of IDS with a more holistic approach to the register. For example, we pointed out above that there were alternative explanations for vowel space expansion, including speech rate (which enhances the perception of emotion; Panneton, Kitamura, Mattock & Burnham, 2006), and smiling (clearly, a communicative signal). Since the present study was designed to test lexically driven enhancement, the ensuing data cannot adjudicate between those other explanations. Nonetheless, we hope to inspire other researchers to pursue at least some of these explanations in future work. Even smiling could be a strong mediating factor, since it appears to capture some otherwise puzzling findings. Let us take, for example, the report that vowel spaces are enlarged in speech to infants, but not pets (Burnham *et al.*, 2003). Although the immediate interpretation is that infants profit from hyperarticulation to learn language, whereas pets do not, this explanation cannot account for the fact that vowels spaces are in fact larger in speech to PUPPIES than in

https://doi.org/10.1017/S0305000914000105 Published online by Cambridge University Press

ADS (Kim, Diehl, Panneton & Moon, 2006). While this result is surprising within a view where promoting language acquisition is the primary factor shaping IDS, it is only natural where smiling, or more generally positive vocal emotion, is a prevalent factor. Naturally, it will be challenging to design studies in which smiling is controlled for, but video-recordings could be used to assess it post hoc.

Our results also suggest one additional feature of IDS that has never been explicitly investigated before. The comparison between the results of the separation and the D(a) suggest that within-category variance is considerably larger in IDS than in ADS. This finding would appear counter-intuitive within language-promoting views of IDS. That is, even if we accept that talkers do not implicitly try to provide infants with better input, why is it that caregivers provide significantly worse input?

It is possible that even if variability is initially detrimental, it eventually leads to better generalization and abstraction (e.g. Homa & Chambliss, 1975; Markman & Maddox, 2003). In fact, Kuhl *et al.* (1997) also observed somewhat greater variability, and specifically hypothesized that stretched vowel spaces allow caregivers to produce more variable vowels, and thus improve category robustness, without increasing the amount of overlap between the categories (p. 685). This appeared to be the case of [eɪ–ɛ] in speech to eleven-month-olds, where there was an increase in between-category distance accompanied by a comparable increase in within-category variability, resulting in overall similarly contrastive categories in both registers. However, we have also shown that the increase in variability can occur in the absence of significant stretching of the distances separating phonological categories. This was the case of the [i–ɪ] contrast in both age groups. Such differences across categories may partially account for why some computer learning studies report improved performance with IDS training (Boer & Kuhl, 2003), others deteriorated performance (Kirchhoff & Schimmel, 2005), and still others no difference (Gauthier & Shi, 2011).

An increase in acoustic variability in IDS such as that recorded here is, in retrospect, evident in much previous work (see figures in Englund & Behne, 2005; Kirchhoff & Schimmel, 2005; Kuhl *et al.*, 1997; Sundberg & Lacerda, 1999). This is a matter that deserves a more thorough evaluation, through experiments specifically designed to assess how much larger the variability is, how much more overlap between categories infants face, whether this differs across different categories and contrasts, and why it happens in the first place. Future work should also evaluate the effects of this increased variability, bearing in mind their likely etiology. For instance, variability in formant values in vowels could be a side effect of fundamental frequency variability, which has been found to improve infants' vowel discrimination (Trainor & Desjardins, 2002) as well as non-linguistic processing (Kaplan, Goldstein, Huckeby & Cooper, 1995; Kaplan & Owren, 1994). We hope future work will shed light on this understudied but, apparently, widespread feature of IDS, whose sources and consequences for language acquisition remain as of yet unexplored.

Although they were not central to our investigation, we would like to underline a few methodological points that bear on IDS research in general. First, this study attempted to capture IDS in a large group of participants. A total of forty-six caregivers of a homogeneous population were recorded, each for 30–45 mins, and analyses were only carried out over caregivers who had considerable numbers of tokens per category. To our knowledge, this is one of the largest sample sizes that has been studied using fine acoustic analyses. Large corpora are desirable in a field where replication is not always straightforward (see Cristia, 2010, for a recent discussion of small sample sizes in IDS research).

The additional effort in gathering such a large pool of data naturally limited the time and resources that could be devoted to transcribing and tagging the speech. In general, it was impossible to control for variables that are often considered in adult research, such as position of the target words within the utterance or utterance length. Clearly, differences in speech rate, duration, positions of the targets within utterances, length of utterances, use of focus, and other such variables are all factors that affect phonetic instantiation. However, these factors have not been controlled in IDS research (Burnham *et al.*, 2003; Cristia, 2010; Englund & Behne, 2005, 2006; Kim *et al.*, 2006; Kitamura *et al.*, 2002; Kondaurova *et al.*, 2012; Kuhl *et al.*, 1997; Lam & Kitamura, 2010, 2012; Liu *et al.*, 2003, 2007, 2009); nor were they controlled here. It could be argued that this is not theoretically problematic, because IDS research seeks to capture the natural input to language acquisition, with all of these factors varying and without assuming that infants normalize or undo their effect real-time just like adults do. Nonetheless, it remains desirable that future work would tease apart some of these variables and assess their potential interactions with register.

In addition, when designing the stimuli, we also attempted to match pairs of lexical items in phonological and lexical variables. The match was not always perfect; for example *pencil* was paired with *best*, where the consonants preceding the target vowel differ in voicing, and which have different numbers of syllables. Mismatches are also evident in previous vowel space research (e.g. *bea**d***, ***p**ot*, *boot*, in the classic Kuhl *et al.*, 1997; *shee**p***, *shoe _sha**rk*** in Burnham *et al.*, 2003; Uther *et al.*, 2007). This is another aspect that could be improved upon in follow-up research.

Finally, much of the evidence for hyperarticulation is based on a single measure (e.g. vowel space size in all the above-mentioned work on vowels; or for consonants a single distance measure between [s-ʃ] in Cristia, 2010; an exception is Liu *et al.*, 2007, who investigated all four Mandarin tones).

Using a single measure could lead to misconceptions, as one does not know how stable and generalizable results are. In the present study, we could assess whether results were stable because we tested two pairs of vowels for each contrast. By including *both* [æ–æ̃] *and* [ε–ε̃], we could conclude that the trend for deterioration found in [æ–æ̃] in 11-month-olds was not a feature of allophonic contrasts, as it was not replicated in [ε–ε̃]. We have further ensured that results were robust to arbitrary analytical decisions. To ease the reader's task, we have focused here on one specific analysis; however, we have carried out many additional ones (e.g. using more or fewer tokens; using non-parametric measures of central tendency and variance; using Mahalanobis distances; separating each acoustic dimension; not normalizing; etc.), all of which revealed the same pattern of results. These supplementary analyses are available on the project website for those readers who are interested in more specific details.

In summary, IDS is likely to be an important source of information for the young language learner. In this article, we have contributed a key piece of evidence on the nature of IDS, by showing that even in a population of talkers where point vowels are hyperarticulated, this enhancement may not occur in other phonemic contrasts. Furthermore, measures of distance that take into account within-category variability suggest that contrasts sometimes deteriorate in speech to infants, likely due to an increase in within-category variability. We concluded that these results, in conjunction with other recent work, invite a revision of the widespread assumption that IDS is primarily shaped by the talkers' goal of promoting language acquisition, which should be replaced with a more holistic, and dyadic, view of the register.

REFERENCES

Bernstein Ratner, N. (1984). Patterns of vowel modification in mother–child speech. *Journal of Child Language* **11**, 557–78.

Boer, B. de & Kuhl, P. K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online* **4**, 2238–46.

Boersma, P. & Weenink, D. (2005). Praat: Doing phonetics by computer (version 5.0.09) [Computer program]. Available from //www.praat.org/.

Burnham, D., Kitamura, C. & Vollmer-Conna, U. (2003). What's new, pussycat? On talking to babies and animals. *Science* **296**, 1435.

Chen, M. (1997). Acoustic correlates of English and French nasalized vowels. *Journal of the Acoustical Society of America* **102**, 2360–70.

Cho, T. (2005). Prosodic strengthening and featural enhancement: evidence from acoustic and articulatory realizations of /a,i/ in English. *Journal of the Acoustical Society of America* **11**, 3867–78.

Cristia, A. (2010). Phonetic enhancement of sibilants in infant-directed speech. *Journal of the Acoustical Society of America* **128**, 424–34.

Cristia, A. (2011). *NSF 0843959: Acoustic cues to allophony*. Available from: http://sites.google.com/site/acrsta/Home/nsf_allophones_corpora.

Cristia, A. (in press). Input to language: the phonetics and perception of infant-directed speech. *Language and Linguistics Compass*. (The annotated database with all IDS studies can be downloaded from https://sites.google.com/site/acrsta/ids_meta-analysis).

Dupoux, E., Peperkamp, S. & Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: stress 'deafness' in simultaneous French–Spanish bilinguals. *Cognition* **114**, 266–75.

Englund, K. & Behne, D. (2005). Infant directed speech in natural interaction: Norwegian vowel quantity and quality. *Journal of Psycholinguistic Research* **34**, 259–80.

Englund, K. & Behne, D. (2006). Changes in infant directed speech in the first six months. *Infant and Child Development* **15**, 139–60.

Escudero, P., Boersma, P., Rauber, A. S. & Bion, R. A. H. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of the Acoustical Society of America* **126**, 1379–93.

Fagel, S. (2010). Effects of smiling on articulation: lips, larynx and acoustics. In A. Esposito, N. Campbell, C. Vogel, A. Hussain & A. Nijholt (eds.), *Development of multimodal interfaces: active listening and synchrony*, 249–303. Berlin/Heidelberg: Springer.

Fernald, A. (2000). Speech to infants as hyperspeech: knowledge-driven processes in early word recognition. *Phonetica* **57**, 242–54.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B. & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language* **16**, 977–1001.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics* **29**, 109–135.

Gauthier, B. & Shi, R. (2011). A connectionist study on the role of pitch in infant-directed speech. *Journal of the Acoustical Society of America* **130**, EL380–EL386.

Gendrot, C. & Adda-Decker, M. (2007). Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrucken, Germany, 1417–20.

Gendrot, C. & Gerdes, K. (2009). Prosodic hierarchy and spectral realization of vowels in French. In H.-Y. Yoo and E. Delais-Roussarie (eds.), *Actes d'IDP*. Available from: http://makino.linguist.jussieu.fr/idp09/actesfr.html.

Hillenbrand, J. M., Clark, M. J. & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America* **109**, 748–63.

Hillenbrand, J. M., Getty, L. A., Clark, M. J. & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* **97**, 3099–3111.

Homa, D. & Chambliss, D. (1975). The relative contributions of common and distinctive information on the abstraction from ill-defined categories. *Journal of Experimental Psychology: Human Learning and Memory* **1**, 351–59.

Kang, K. & Guion, S. G. (2008). Clear speech production of Korean stops: changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America* **124**, 3909–917.

Kaplan, P. S., Goldstein, M. H., Huckeby, E. R. & Cooper, R. P. (1995). Habituation, sensitization, and infants' responses to motherese speech. *Developmental Psychobiology* **28**, 45–57.

Kaplan, P. S. & Owren, M. J. (1994). Dishabituation of visual attention in 4-month-olds by infant-directed frequency sweeps. *Infant Behavior and Development* **17**, 347–58.

Keating, P., Cho, T., Fougeron, C. & Hsu, C. S. (2003). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden & R. Temple (eds.), *Papers in laboratory phonology VI*, 143–61. Cambridge: Cambridge University Press.

Kim, H. J., Diehl, M., Panneton, R. & Moon, C. (2006). Hyperarticulation in mothers' speech to babies and puppies. Poster presented at the International Conference on Infant Studies, Kyoto, Japan.

Kirchhoff, K. & Schimmel, S. (2005). Statistical properties of infant-directed versus adult-directed speech: insights from speech recognition. *Journal of the Acoustical Society of America* **117**, 2238–46.

Kitamura, C. & Burnham, D. (2003). Pitch and communicative intent in mothers' speech: adjustments for age and sex in the first year. *Infancy* **4**, 85–110.

Kitamura, C., Thanavishuth, C., Burnham, D. & Luksaneeyanawin, S. (2002). Universality and specificity in infant-directed speech: pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development* **24**, 372–92.

Kondaurova, M. V., Bergeson, T. R. & Dilley, L. (2012). Effects of deafness on acoustic characteristics of American English tense/lax vowels in maternal speech to infants. *Journal of the Acoustical Society of America* **132**, 1039–49.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U. & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* **277**, 684–86.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M. & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B* **363**, 979–1000.

Lam, C. & Kitamura, C. (2010). Maternal interactions with a hearing and hearing-impaired twin: similarities and differences in speech input, interaction quality, and word production. *Journal of Speech, Language, and Hearing Research* **53**, 543–55.

Lam, C. & Kitamura, C. (2012). Mommy, speak clearly: induced hearing loss shapes vowel hyperarticulation. *Developmental Science* **15**, 212–21.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* **35**, 1773–81.

Liu, H. M., Kuhl, P. K. & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science* **6**, F1–F10.

Liu, H. M., Tsao, F. M. & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology* **43**, 912–17.

Liu, H. M., Tsao, F. M. & Kuhl, P. K. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: a longitudinal study. *Journal of Child Language* **36**, 909–922.

Lively, S. E., Logan, J. S. & Pisoni, D. E. (1993). Training Japanese listeners to identify /r/ and /l/ ii. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* **94**, 1242–55.

Malsheen, B. J. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (eds.), *Child phonology Vol. 2*, 173–84. San Diego, CA: Academic Press.

Markman, A. B. & Maddox, W. T. (2003). Classification of exemplars with single- and multiple-feature manifestations: the effects of relevant dimension variation and category structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **29**, 107–117.

Newman, R. S., Clouse, S. A. & Burnham, J. (2001). The perceptual consequences of acoustic variability in fricative production within and across talkers. *Journal of the Acoustical Society of America* **109**, 1181–96.

Panneton, R., Kitamura, C., Mattock, K. & Burnham, D. (2006). Slow speech enhances younger but not older infants' perception of vocal emotion. *Research in Human Development* **3**, 7–19.

Pruthi, T. (2007). Analysis, vocal-tract modeling and automatic detection of vowel nasalization. *Unpublished doctoral dissertation*, University of Maryland.

Rost, G. C. & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science* **2**, 339–49.

Snow, C. E. & Ferguson, C. A. (1977). *Talking to children: language input and acquisition*. Cambridge: Cambridge University Press.

Soderstrom, M. (2007). Beyond babytalk: re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review* **27**, 501–532.

https://doi.org/10.1017/S0305000914000105 Published online by Cambridge University Press

Sundberg, U. & Lacerda, F. (1999). Voice Onset Time in speech to infants and adults. *Phonetica* **56**, 186–99.

Tartter, V. (1980). Happy talk: perceptual and acoustic effects of smiling on speech. *Perception and Psychophysics* **27**, 24–27.

Trainor, L. J. & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin and Review* **9**, 335–40.

Uther, M., Knoll, M. A. & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication* **49**, 2–7.

van de Weijer, J. (1998). *Language input for word discovery* (Max Planck Series in Psycholinguistics 9). Nijmegen: Max Planck.