Received: 12 March 2023 | Revised: 8 August 2023 | Accepted: 10 August 2023

DOI: 10.1111/desc.13442

Developmental Science WILEY

SHORT REPORT

Relating referential clarity and phonetic clarity in infant-directed speech

Caroline Beech | Daniel Swingley

Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Correspondence
Caroline Beech, Department of Psychology, University of Pennsylvania, 425 S University Ave, Philadelphia, PA, 19104, USA.
Email: cbeech@sas.upenn.edu

Funding information
National Institutes of Health; National Science Foundation

Abstract

Psycholinguistic research on children's early language environments has revealed many potential challenges for language acquisition. One is that in many cases, referents of linguistic expressions are hard to identify without prior knowledge of the language. Likewise, the speech signal itself varies substantially in clarity, with some productions being very clear, and others being phonetically reduced, even to the point of uninterpretability. In this study, we sought to better characterize the language-learning environment of American English-learning toddlers by testing how well phonetic clarity and referential clarity align in infant-directed speech. Using an existing Human Simulation Paradigm (HSP) corpus with referential transparency measurements and adding new measures of phonetic clarity, we found that the phonetic clarity of words' first mentions significantly predicted referential clarity (how easy it was to guess the intended referent from visual information alone) at that moment. Thus, when parents' speech was especially clear, the referential semantics were also clearer. This suggests that young children could use the phonetics of speech to identify globally valuable instances that support better referential hypotheses, by homing in on clearer instances and filtering out less-clear ones. Such multimodal "gems" offer special opportunities for early word learning.

KEYWORDS
Human Simulation Paradigm, infant-directed speech, language acquisition, phonetic clarity, referential clarity, word learning

Research Highlights

- In parent-infant interaction, parents' referential intentions are sometimes clear and sometimes unclear; likewise, parents' pronunciation is sometimes clear and sometimes quite difficult to understand.
- We find that clearer referential instances go along with clearer phonetic instances, more so than expected by chance.
- Thus, there are globally valuable instances ("gems") from which children could learn about words' pronunciations and words' meanings at the same time.
- Homing in on clear phonetic instances and filtering out less-clear ones would help children identify these multimodal "gems" during word learning.

1 | INTRODUCTION

To infants, much of the language they hear starts out as a mystery. At first, infants know nothing about the words that speech contains, or which ideas those words represent. Even the notion that speech refers to things in the world may be a discovery; infants may begin by thinking of speech primarily as an affective signal (Fernald, 1989; though see Ferry et al., 2010). Similarly, infants must discover that utterances are made up of smaller units—words—that are reused by speakers and that convey a similar concept each time. Somehow, given enough experience with language, children eventually solve some of these mysteries; and even in the first year, they learn the spoken form and the meaning of a variety of words (e.g., Bergelson & Swingley, 2012; Hallé & Boysson-Bardies, 1994; Jusczyk & Aslin, 1995; Swingley, 2005).

To understand how children succeed at this task, we must try to characterize what language is like from the child's perspective. How much information and what kind of information exists in the input to the language-learning process? By analyzing corpora of parents' speech to children, researchers have identified many potential challenges presented by the linguistic environment (e.g., Bergelson et al., 2019; Cristia & Seidl, 2014; Höhle et al., 2004). In particular, children have to contend with rampant ambiguity both in the phonetic forms with which words are realized, and in the referential intentions of the speaker.

The phonetic clarity and referential clarity of speech are highly variable. Sometimes the spoken signal is an inscrutable muddle, and sometimes the speaker's intentions are murky. But not always. On occasion, the spoken signal is particularly clear, and in some cases the reference relatively transparent. In the present work, we consider these instances, and ask: do the clearer phonetic realizations and the clearer referential contexts tend to go together, in single, multimodally transparent conversational events? Or are these dually helpful events no more common than would be expected by chance? If such instances represent especially informative and important learning events, it is useful to know how frequently they occur.

Considering referential ambiguity: without any prior constraints, there are many possible referents or meanings that children might consider upon hearing an unknown word. For example, the word might refer to an object, a property or part of an object, an event, an abstract notion with no sensory correlate, or a combination of these (e.g., Quine, 1960). If we assume that a word refers to something concrete, the same scene can still invite multiple perspectives or contain many copresent objects or actions, which may or may not even include the intended referent. Clearly, correctly identifying referents requires some convergence between the child and the parent, and this seems easier in some cases than in others (Gleitman et al., 2005).

To assess the referential clarity of speech to a learner with fairly sophisticated concepts but without access to much linguistic information (their view of the child in the early stages of word learning), Gillette et al. (1999) developed the Human Simulation Paradigm (HSP). In the HSP, adult participants watch videos of parents talking to their toddlers. For the participants, the sound is turned off, except for a beep or nonsense word signaling the onset of a particular target word in the parent's speech. Then, participants are asked to guess the meaning of the word that parents said at that moment, using visual observation alone, like a young child who lacks access to additional linguistic cues. Studies using this method, and related analyses of children's language environments, show that in at least some cases, the nature of parent-child interaction can help to constrain the set of solutions to this problem. For example, the referent of an object-labeling event is frequently the dominant object in the child's visual field (Pereira et al., 2014). In other cases, though, the referential ambiguity seems profound (e.g., Cartmill et al., 2013; Trueswell et al., 2016). The present study does not try to evaluate which of these scenarios predominates in the experience of language learners. Instead, we introduce the question of whether referential clarity is related to phonetic clarity.

When adults listen to speech, we leverage our knowledge of the surrounding linguistic context to help identify individual words. This allows us to contend with widespread variability in how words can be pronounced, even by a single speaker. Studies of phonetic assimilation (e.g., Buckler et al., 2018; Dilley et al., 2014) and reduction processes (e.g., Lahey & Ernestus, 2014; Shockey & Bond, 1980) reveal that such variability is prevalent in speech to children too. As a result, when the linguistic context is removed, mirroring the infant's lack of knowledge, individual words of infant-directed speech are frequently unintelligible to adult listeners, perhaps even more so than isolated words of adult-directed speech are (Bard & Anderson, 1983). Yet, like referential clarity, the phonetic clarity of infant-directed speech is variable. While there are many uninterpretable pronunciations, there are also some instances with little coarticulation or reduction, where the segments of the word are distinctly realized (Cychosz et al., 2021).

In the present study, our goal was to investigate how much the phonetically clear cases and the referentially clear cases coincide in infant-directed speech. To do this, we used an existing audiovisual corpus of parents talking to their toddlers, and tested whether the phonetic clarity with which a given word was produced (judged by adult listeners) predicted the referential clarity at that moment (assessed using HSP). If phonetic clarity predicts referential clarity, it would suggest that from the child's perspective, early word learning involves homing in on globally valuable instances, word-learning gems in a slurry of phonetic and referential ambiguity. If, on the other hand, referential clarity is independent of phonetic clarity in the input, it would imply that children must integrate phonetically clear tokens with separate referentially clear instances that exhibit less phonetic clarity. Thus, answering this question serves to direct theories of word learning more generally.

Building on previous work investigating repetition effects (e.g., Bard et al., 2000; Fowler & Housum, 1987; Lam & Watson, 2010; Pate & Goldwater, 2011), we also explored how the phonetic clarity of a word varied over a discourse and whether this was related to referential clarity. Finally, we analyzed the acoustics of the words in our dataset to better understand the lower level properties underlying our measures of phonetic clarity.

Our focus here is on how the word-form:referent pairs could be learned for children's earliest words. Language learning as a whole

is of course more complex than this mapping problem. Thus, even if word-learning gems exist for certain concrete words, this is not to say that these same moments are equally useful for learning about other aspects of the language (e.g., question formation, negation, sentence structure). Instead, word-learning gems, as we use the term, are moments that elucidate the phonetic form and the referent/meaning of a particular word simultaneously, which could be helpful in the initial stages of language learning.

2 | METHODS

2.1 | Corpus

We used an existing corpus of 560 forty-second video clips ("vignettes") of parents speaking to their 14- or 18-month-old in the home (Cartmill et al., 2013). The vignettes came from 56 families (10 vignettes each) of typically developing, monolingual English-learning children, who were recorded as part of a larger longitudinal study of language development (Goldin-Meadow et al., 2014). Cartmill et al. (2013) extracted the vignettes from longer recordings such that each vignette represented a randomly selected instance of one of the 10 most common concrete nouns produced by that parent. Thus, the contexts in the vignettes varied and were not limited to object play or moments when the target word's referent was present.

Each vignette in the corpus also had an associated measure of referential clarity collected using the HSP (introduced above). This measure, *HSP accuracy*, ranges from 0 to 1 and reflects the proportion of English-speaking adults who correctly guessed parents' referential intention (the target word) after watching the muted video, with one or more beeps indicating the onset(s) of the target word in the parent's speech.

Following our preregistration, we analyzed a subset of these vignettes from 27 unique families and 10 unique target words (ball, bear, book, dog, eye, hair, hand, kiss, nose, and shoe). The original dataset included vignettes from 56 families, but because the vignettes were selected from instances of that family's top 10 concrete nouns, not all of the families contributed data for the same words. Indeed, some words had very little data (just one or two families). In addition, high-informative vignettes (HSP accuracy ≥0.5) were rare across the corpus (only 12.5% of vignettes). In deciding which audio to transcribe and segment, we endeavored to create a more balanced dataset, with a higher proportion of high-informative vignettes. In particular, we selected families and words for our analysis such that each included family and each included word had at least two "high-informative" vignettes. In addition, each included word was required to have at least one vignette from each level of informativity—"low-informative" (HSP accuracy ≤0.1), "medium-informative" (HSP accuracy between 0.1 and 0.5), and "high-informative" (HSP accuracy ≥0.5). Of the 177 vignettes that met these initial criteria, seven were excluded because the speech was completely unintelligible to trained coders, and two were excluded because of idiomatic usage (e.g., "keep an *eye* on this"), leaving us with 168 vignettes in total.

2.2 | Measures of phonetic clarity

In order to collect new measures of phonetic clarity for the target words, we first separated the original audio (i.e., parents' speech, not beeps) from the videos. From these audio files, trained coders isolated each instance of the target word in each file using Praat (Boersma & Weenink, 2022). As in previous studies using this corpus, morphological variants such as "dogs" or "doggy" for "dog" were counted as instances of the target word. These isolated words composed the sample of words to be evaluated.

We presented the words to adult English speakers, and asked them to judge the clarity of the speech on a scale from 1 to 5, and to transcribe the word they thought was said. The presentation order was quasi-randomized to vary across listeners while ensuring that no listener ever heard the same target word on two consecutive trials. Before starting the task, listeners were explicitly instructed to make their judgments based only on the clarity of the speech rather than on properties of the recording such as background noise. This was intended to ensure that listeners distinguished phonetic clarity, a property of the speaker's articulation of the word, from recording quality, or the presence of nuisance features like low volume, background noise, or noisy single events such as door-closings. In case listeners were not able to assess phonetic clarity independently of recording quality, however, the same listeners also made recording quality judgments (scaled from 1 to 5) for each 40-s file in a separate task, based on a 2-s snippet that preceded one of the target-word instances and that did not contain another instance of the target word. Before starting this task, listeners were explicitly instructed to "make [their] judgments based only on the quality of the recording (e.g., mechanical noise, TV drowning out speech), not on whether the snippet contains a complete sentence" and were presented with three example snippets of good, average, and bad recording quality from a separate corpus.

Because the Cartmill et al. (2013) audio data are protected and can only be listened to by the original research team and their collaborators, listeners were recruited from among our lab personnel. The only change we made compared to our preregistration was to collect judgments and transcriptions from seven listeners instead of five, reflecting the slightly larger size of the lab at the time of measurement. None of the judges was involved in extracting the audio samples from the corpus, or had any knowledge of the vignettes' HSP accuracy.

For each isolated word token, the listening task data provided three measures of phonetic clarity: *phonetic clarity rating* (the average of listeners' clarity judgments), *transcription accuracy* (the proportion of listeners who could correctly identify the word), and *transcription distance* (the average phonological distance between listeners' guesses and the word). As the metric for phonological distance, we used normalized Levenshtein distance (number of phonemes changed / maximum possible changes), which ranges from 0 (identical) to 1 (no phonemes in common). While not part of our preregistration, transcription distance affords a more fine-grained measure than transcription accuracy because it distinguishes between tokens that can be tran-

TABLE 1 Results of a mixed effects logistic regression model predicting referential transparency (Human Simulation Paradigm [HSP] accuracy) from phonetic clarity rating with random intercepts for word and family.

Coefficient	Beta	SE	*p*
Intercept	−0.592	0.356	0.003*
Phonetic clarity rating	0.192	0.165	0.043*
Number of mentions	0.016	0.100	0.734
Position in utterance (final vs. medial)	−0.107	0.413	0.663
Position in utterance (initial vs. medial)	−0.151	0.677	0.857
Number of obs: 168. Groups: word (10), family (27). *p*-values calculated using bootstrap resampling.			
Group	Variance		
Word (intercept)	0.009		
Family (intercept)	0		

TABLE 2 Results of a mixed effects logistic regression model predicting referential transparency (Human Simulation Paradigm [HSP] accuracy) from transcription accuracy with random intercepts for word and family.

Coefficient	Beta	SE	*p*
Intercept	−0.601	0.359	0.001*
Transcription accuracy	0.462	0.495	0.106
Number of mentions	0.036	0.098	0.474
Position in utterance (final vs. medial)	−0.098	0.417	0.678
Position in utterance (initial vs. medial)	−0.088	0.672	0.974
Number of obs: 168. Groups: word (10), family (27). *p*-values calculated using bootstrap resampling.			
Group	Variance		
Word (intercept)	0.017		
Family (intercept)	0		

scribed fairly faithfully but not perfectly and tokens that cannot be transcribed with any fidelity.

As described in our preregistration, we corrected these measures for recording quality where necessary: When the raw phonetic clarity measurements were significantly correlated with the average recording quality judgments (which was the case for the phonetic clarity rating and transcription accuracy), we applied a simple linear regression predicting phonetic clarity from recording quality and extracted the residuals ([raw phonetic clarity measurement] - [expected value given recording quality]) to use instead of the raw measurements in subsequent analyses.

2.3 | Other measures

In addition to our human measures of phonetic clarity, we also collected a few automated measurements. These consisted of *duration* (how long was the target word, corrected for its expected length given the phonemes in the word's canonical pronunciation), *mean pitch* (calculated using Praat, and centered and scaled for each talker), *position in utterance* (computed from trained coders' transcriptions), and *number of repetitions* within the 40-second file.

TABLE 3 Results of a mixed effects logistic regression model predicting referential transparency (Human Simulation Paradigm [HSP] accuracy) from transcription distance with random intercepts for word and family.

Coefficient	Beta	SE	*p*
Intercept	−0.594	0.357	0.002*
Transcription distance	−0.715	0.560	0.022*
Number of mentions	0.020	0.099	0.682
Position in utterance (final vs. medial)	−0.113	0.415	0.645
Position in utterance (initial vs. medial)	−0.103	0.673	0.943
Number of obs: 168. Groups: word (10), family (27). *p*-values calculated using bootstrap resampling.			
Group	Variance		
Word (intercept)	0.017		
Family (intercept)	0		

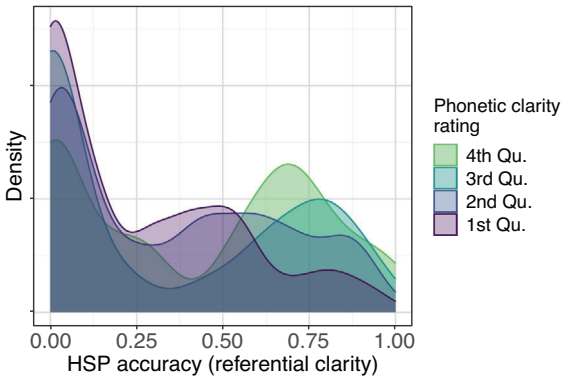

FIGURE 1 Distribution of Human Simulation Paradigm (HSP) accuracy by phonetic clarity rating. When parents produced a word less clearly (phonetic clarity rating in the 1st quartile, shown in purple), the intended referent or meaning was usually hard to guess from the visual scene alone (low HSP accuracy, probability mass toward left side of the plot). By contrast, the phonetically clearest instances (phonetic clarity rating in the 4th quartile, shown in green) were more likely to have referents that were clear from the visual context (high HSP accuracy, to the right side of the plot).

3 | RESULTS

For each of our three phonetic clarity measures, we asked how well these measures predicted HSP accuracy, that is, whether clearer productions in parents' speech are more likely to coincide with transparent references. Because HSP accuracy is a proportion, we used mixed effects logistic regression models to predict HSP accuracy from phonetic clarity with random intercepts for word and family. Following

TABLE 4 Results of a mixed effects logistic regression model predicting referential transparency (Human Simulation Paradigm [HSP] accuracy) from the difference in phonetic clarity rating (first—later mentions), with random intercepts for word and family. There were 90 vignettes with multiple mentions, excluding vignettes in which the child or another speaker produced the target word.

Coefficient	Beta	SE	*p*
Intercept	−0.751	0.264	<.001*
Difference in phonetic clarity rating (first - later mentions)	0.112	0.283	0.590
Phonetic clarity rating of first mention	0.378	0.271	0.069
Number of obs: 90. Groups: word (10), family (27). *p*-values calculated using bootstrap resampling.			
Group	Variance		
Word (intercept)	0		
Family (intercept)	0		

our preregistration, we also included fixed effects for number of repetitions and position in utterance to test whether more repetitions or privileged utterance positions were associated with higher HSP accuracy, all else being equal, although neither of these variables was ultimately significant. Tables 1–3 summarize the primary results. All analyses were conducted in R (R Core Team, 2019) and *p*-values were calculated using the glmboot package (Humphrey, 2022) for bootstrap resampling.

Using the phonetic clarity rating measure, we found that better phonetic clarity of the first instance of the target word significantly predicted higher HSP accuracy (Table 1; Figure 1). In other words, when parents' speech was especially clear (as measured from just the audio), the referential semantics were more transparent (as measured from just the video). Unsurprisingly, this relationship was not as strong using the more coarse-grained measure of transcription accuracy (Table 2), but using the more fine-grained transcription distance measure, we found that smaller transcription distance (greater proximity of the transcription of the first instance and the correct form) significantly predicted higher HSP accuracy (Table 3).

These results suggest that rather than trading off across conversational events, phonetic and referential clarity tend to go together. Thus, filtering out phonetically less clear productions might actually leave children in a better position to learn a word's meaning, allowing them to home in on multimodally transparent instances.

3.1 | Clarity across the discourse

Since many of the vignettes we considered (106 out of 168) contained multiple instances of the target word, we also examined how the *average* phonetic clarity and *maximum* phonetic clarity across repetitions of the target word related to HSP accuracy. When considering judgments of every instance of the target word in a vignette, rather than only the first instance, the relationship to referential clarity was not, or only marginally, significant (see Appendix, Tables S1–S6). This could imply that first mentions are especially important, or that the phonetic clarity of word repetitions is driven by a different process than the one that usually causes phonetic clarity and referential clarity to align.

In addition, we used the vignettes with multiple mentions to investigate changes in clarity across the discourse. We focus on *phonetic clarity rating* for simplicity. Figure 2 shows the clarity of first mention compared to subsequent repetitions, excluding vignettes in which the child or another speaker produced the target word. On average, we found that the phonetic clarity of later mentions was significantly lower

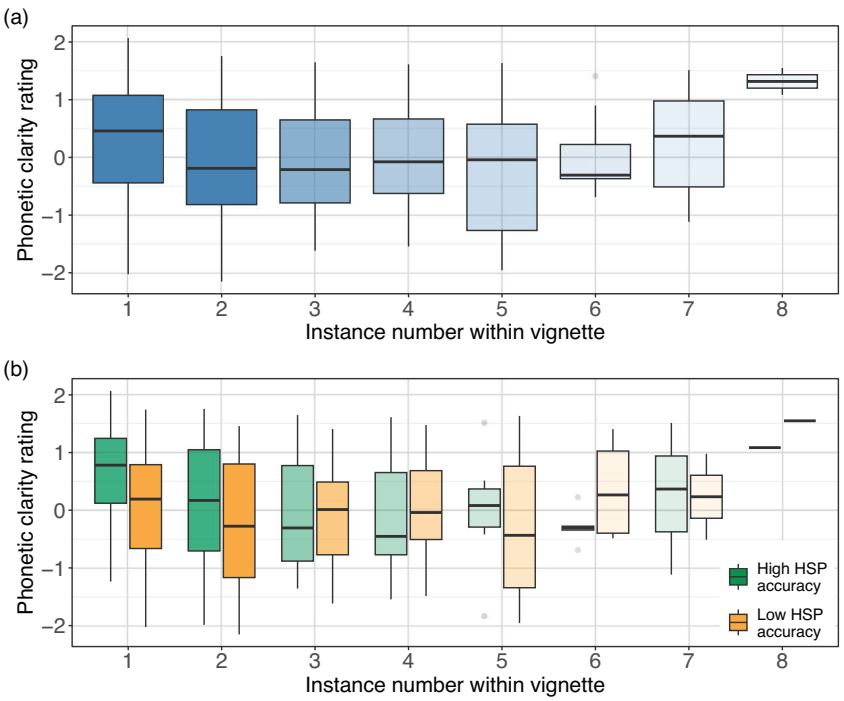

FIGURE 2 Phonetic clarity across instances. Panel A: Boxplots show the phonetic clarity rating (corrected for recording quality) of first and later mentions, with more transparent boxplots indicating fewer observations. Only vignettes with at least two mentions of the target word, excluding those in which the child or another speaker produced the target word, were considered. Panel B shows the same data split by referential clarity (Human Simulation Paradigm [HSP] accuracy above or below 50%). Later mentions were usually less clear than first mentions for both high and low HSP accuracy vignettes.

than the phonetic clarity of first mentions ($t(89) = 4.64$, $p < 0.001$), in line with previous work in this area (e.g., Fowler & Housum, 1987). However, the difference in clarity between the first mention and the later mentions was not predictive of HSP accuracy after controlling for the clarity of the first mention (Table 4). Later mentions tended to be less clear than first mentions regardless of referential transparency (Figure 2 panel B).

3.2 | Acoustic properties

To examine which acoustic properties might underlie our holistic human measures of phonetic clarity, we conducted simple linear regressions with phonetic clarity as the outcome, and duration and mean pitch, each plausibly associated with hyperarticulation, as the two predictors. As mentioned in the previous section, we corrected duration for the expected length of the word given the phonemes in its canonical pronunciation and the average length of those phonemes in a separate corpus: duration = log (raw duration / sum [average duration of each phone in the infant-directed speech of several talkers]). Mean pitch values were centered and scaled (divided by their standard deviation) within each talker. We found that duration significantly predicted all three phonetic clarity measures—phonetic clarity rating ($\beta = 0.76$, SE = 0.11, $p < 0.001$), transcription accuracy ($\beta = 0.16$, SE = 0.04, $p < 0.001$), and transcription distance ($\beta = -0.12$, SE = 0.03, $p < 0.001$)—in the expected direction. Mean pitch, on the other hand, was not significantly related to transcription accuracy ($\beta = 0.01$, SE = 0.02, $p = 0.418$) or transcription distance ($\beta = -0.01$, SE = 0.02, $p = 0.390$), and only marginally predictive of phonetic clarity rating ($\beta = 0.09$, SE = 0.05, $p = 0.077$) after controlling for duration. When duration was not held constant, the relationship between pitch and phonetic clarity rating attained significance ($\beta = 0.10$, SE = 0.05, $p = 0.047$), with higher pitch predicting higher phonetic clarity rating.

4 | CONCLUSIONS

In this study, we investigated the relationship between phonetic and referential clarity in infant-directed speech. Using an existing HSP corpus, we found that the phonetic clarity of first mentions (*phonetic clarity rating* or *transcription distance*) significantly predicted the referential clarity of the scene (*HSP accuracy*, i.e., how easy it was to guess the referent from visual information alone). Thus, clearer productions in parents' speech were more likely to coincide with transparent references. This suggests that word learning could involve homing in on multimodally transparent instances, in which both the speech and the meaning are especially clear.

The fact that transcription distance predicted referential transparency better than transcription accuracy did suggests that some instances of words that are too hypoarticulated to be recognized reliably, but are phonologically close, still participate in the association between spoken clarity and referential transparency. It is not known how infants interpret such situations. For example, if a parental realization of a word contained an ambiguous sound (like "bear" with a hard-to-categorize /b/), or could be interpreted as having a distinct pronunciation (like "bear" as "vair"), what would the infant make of it? Infants might be able to store an underspecified or gradient representation of the word (e.g., Vihman et al., 1994; Waterson, 1971), or, they might probabilistically select a categorical representation (e.g., a segmental representation of "bear", of "vair", or of "nair") to store, with the most frequently perceived form eventually winning out in the lexicon (e.g., Ranbom & Connine, 2007).

It is worth considering how children would detect word-learning gems in their speech environment. Previous work by Trueswell et al. (2016) suggests that observers could use certain dynamic visuo-social cues, such as the sudden appearance of an object, to identify rare moments of referential clarity. Our results point to an additional possibility. Since phonetic clarity predicts referential clarity, if infants were to attend more to clearer speech, filtering out less-clear productions, this filtering on the phonetics side would also leave them with a better set of referential hypotheses. While it is certainly not the case that all phonetically clear instances are multimodal gems, on average their referential clarity would be higher than in the unfiltered input.

Eventually, of course, children must learn to make sense of less clear pronunciations too. How might this happen? One possibility is that children might take advantage of the kind of discourse/second mention effects observed in this and previous studies, first recognizing phonetically clear initial instances and then extending this lexical interpretation to subsequent, less clear instances. Ultimately, this question of when and how children come to understand less clear pronunciations is an empirical one, which will need to be answered by future research.

One limitation of the present work is the size of the corpus we considered. Future work using a larger or longitudinal corpus could provide more insight into the generalizability of our findings or the nature of this relationship over the course of development.

This study does not speak to *why* phonetic and referential clarity are related. It could be that this property of the input reflects intentional teaching on the part of parents. This seems unlikely, though, because parents as intentional teachers might also be expected to increase their phonetic clarity when the intended referent is *not* transparent, in compensation; and intentional teachers might not be expected to hold off on phonetic emphasis after the first mention. As an alternative explanation, simultaneous phonetic highlighting and referential transparency could emerge naturally in those parent-child conversations that focus on a particular object as the key element of the discourse.

Previous explorations of phonetic variability and of referential transparency have rarely considered both elements together. The significance of the present result depends to some degree on what children can do with phonetically or referentially obscure instances of words. If such instances are discarded or have minimal influence in learning, a lack of coordination between phonetically and referentially obscure instances would constrict the effective language learning

dataset enormously, as the infant would have to rely on infrequent chance cooccurrences of phonetic and referential clarity, or learn words' phonetic forms and words' meanings from separate instances. We see here that this is not the case: to the extent that our parent-child dyads are representative, it appears that infants can count on some coordination between phonetic and referential clarity. Thus, the phonetics of speech provide an additional cue to the availability of word meanings.

ACKNOWLEDGMENTS

This work received financial support from NSF grant 1917608 to Daniel Swingley, and from NIH (NIDCD) grant T32 DC016903, on which Caroline Beech was a trainee. We would also like to thank John Trueswell for making the audio and HSP accuracy data available to us and providing thoughtful feedback throughout the project.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in OSF at https://doi.org/10.17605/OSF.IO/U7W52.

ORCID

Caroline Beech https://orcid.org/0000-0002-8357-9091

REFERENCES

Bard, E. G., & Anderson, A. H. (1983). The unintelligibility of speech to children. *Journal of Child Language*, *10*(2), 265–292. https://doi.org/10.1017/S0305000900007777

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*(1), 1–22. https://doi.org/10.1006/jmla.1999.2667

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do North American babies hear? A large-scale cross-corpus analysis. *Developmental Science*, *22*(1), e12724. https://doi.org/10.1111/desc.12724

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* (6.2.17).

Buckler, H., Goy, H., & Johnson, E. K. (2018). What infant-directed speech tells us about the development of compensation for assimilation. *Journal of Phonetics*, *66*, 45–62. https://doi.org/10.1016/j.wocn.2017.09.004

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, *41*(4), 913–934. https://doi.org/10.1017/S0305000912000669

Cychosz, M., Edwards, J. R., Bernstein Ratner, N., Torrington Eaton, C., & Newman, R. S. (2021). Acoustic-lexical characteristics of child-directed speech between 7 and 24 months and their impact on toddlers' phonological processing. *Frontiers in Psychology*, *12*, 712647. https://doi.org/10.3389/fpsyg.2021.712647

Dilley, L. C., Millett, A. L., McAuley, J. D., & Bergeson, T. R. (2014). Phonetic variation in consonants. *Journal of Child Language*, *41*(1), 153–173. https://doi.org/10.1017/S0305000912000670

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, *60*(6), 1497–1510. https://doi.org/10.2307/1130938

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development*, *81*(2), 472–479. https://doi.org/10.1111/j.1467-8624.2009.01408.x

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*(5), 489–504. https://doi.org/10.1016/0749-596X(87)90136-7

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176. https://doi.org/10.1016/S0010-0277(99)00036-0

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, *1*(1), 23–64. https://doi.org/10.1207/s15473341lld0101_4

Goldin-Meadow, S., Levine, S. C., Hedges, L. V., Huttenlocher, J., Raudenbush, S. W., & Small, S. L. (2014). New evidence about language and cognitive development based on a longitudinal study: Hypotheses for intervention. *American Psychologist*, *69*(6), 588–599. https://doi.org/10.1037/a0036886

Hallé, P. A., & de Boysson-Bardies, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior and Development*, *17*(2), 119–129. https://doi.org/10.1016/0163-6383(94)90047-7

Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*, *5*(3), 341–353. https://doi.org/10.1207/s15327078in0503_5

Humphrey, C. (2022). *glmmboot: Bootstrap resampling for mixed effects and plain models* (0.6.0). https://github.com/ColmanHumphrey/glmmboot

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23. https://doi.org/10.1006/cogp.1995.1010

Lahey, M., & Ernestus, M. (2014). Pronunciation variation in infant-directed speech: Phonetic reduction of two highly frequent words. *Language Learning and Development*, *10*(4), 308–327. https://doi.org/10.1080/15475441.2013.860813

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory and Cognition*, *38*(8), 1137–1146. https://doi.org/10.3758/MC.38.8.1137

Pate, J. K., & Goldwater, S. (2011). Predictability effects in adult-directed and infant-directed speech: Does the listener matter? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*, 1569–1574. https://escholarship.org/uc/item/2vb2042t

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin and Review*, *21*(1), 178–185. https://doi.org/10.3758/s13423-013-0466-4

Quine, W. V. O. (1960). *Word and object: An inquiry into the linguistic mechanisms of objective reference*. MIT Press.

R Core Team. (2019). *R: A language and environment for statistical computing* (3.6.2). R Foundation for Statistical Computing. https://www.r-project.org/

Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298. https://doi.org/10.1016/j.jml.2007.04.001

Shockey, L., & Bond, Z. S. (1980). Phonological processes in speech addressed to children. *Phonetica*, *37*(4), 267–274. https://doi.org/10.1159/000259996

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, *8*(5), 432–443. https://doi.org/10.1111/j.1467-7687.2005.00432.x

Trueswell, J. C., Lin, Y., Armstrong, B. F., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent-child interactions. *Cognition*, *148*, 117–135. https://doi.org/10.1016/j.cognition.2015.11.002

Vihman, M., Velleman, S., & McCune, L. (1994). How abstract is child phonology? Towards an integration of linguistic and psychological approaches. In M. Yavas (Ed.), *First and second language phonology* (pp. 9–44). Singular Publishing Group.

Waterson, N. (1971). Child phonology: A prosodic view. *Journal of Linguistics*, *7*(2), 179–211. https://doi.org/10.1017/S0022226700002917

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Beech, C., & Swingley, D. (2023). Relating referential clarity and phonetic clarity in infant-directed speech. *Developmental Science*, 1–8. https://doi.org/10.1111/desc.13442