Nature's distributional-learning experiment: Infants' input, infants' perception, and computational modeling

Benders, A.T.

Publication date
2013

6

DISCUSSION AND CONCLUSION: EVALUATING NATURE'S DISTRIBUTIONAL-LEARNING EXPERIMENT

Abstract

This dissertation is the result of an integrated research program to study distributional learning (the acquisition of speech-sound categories on the basis of the shape of the input distribution) in practice: Nature's distributional learning experiment. The input distributions in this experiment were distributions as observed in mother-child interaction (Part I; Chapters 2 and 3). The results of learning from this input were observed in the native-language perception patterns of infants (Part II; Chapters 3 and 4). Computer models with only a distributional-learning device were then trained on the input distribution and were found to account for the infant perception data (Part 3; Chapter 5). These results lend strong support to the distributional-learning hypothesis.

6.1 Summary of the study aims

In this dissertation it was investigated whether infants learn their phoneme categories through distributional learning. To this end, I pursued a three-part research program that I called "nature's distributional-learning experiment":

Part I) investigate the acoustic properties and the auditory distributions of some phonemes in the infants' environment (Chapters 2 and 3);

Part II) investigate infants' perception of those same phonemes (Chapters 3 and 4);

Part III) explain the perception as found in Part II from the distributions found in Part I through computationally simulated distributional learning (Chapter 5).

It was argued in the Introduction (Chapter 1) that similarities and differences between infants' input and perception can be better detected if the input distributions and infants' perception are investigated along multiple auditory dimensions. Because Dutch vowels /ɑ/ and /aː/ as the test case typically differ in both vowel quality and duration, the program was pursued with /ɑ/ and /aː/ as the test case.

6.2 Summary of the empirical results: Similarities between infants' input and perception

In Chapters 2 and 3 it was shown that /ɑ/ and /aː/ in Dutch IDS differed in their mean vowel quality and duration. Nevertheless, the pooled frequency distribution of the vowel quality values of the /ɑ/- and /aː/-tokens was monomodal, as was the pooled distribution of their duration values. The pooled distribution of the two vowels only had separate local maxima for /ɑ/ and /aː/ in a two-dimensional auditory space that was defined by both vowel quality and duration. These results from Chapter 3 suggest that if infants learn the /ɑ/–/aː/ contrast through distributional learning, they must learn it from the two-dimensional frequency distribution. Dutch infants should learn that each vowel is associated with a specific vowel quality as well as with a specific duration.

In the discrimination task in Chapter 3, Dutch 11- and 15-month-old infants were found to discriminate better between /ɑ/ and /aː/ when the difference between the vowels was signalled by both cues than when it was signalled by only vowel quality or only duration. This first perception result shows that Dutch infants know that /ɑ/ and /aː/ typically differ in two cues. In the categorization task in Chapter 4, Dutch 15-month-old infants allocated their attention differently to the atypical vowel sounds [ɑː] and [a] than to vowel sounds with the typical combinations of vowel quality and duration, [ɑ] and [aː]. This second perception result shows again that Dutch infants associate their representations of /ɑ/ and /aː/ with combinations of vowel quality and duration. Moreover, the larger the infant's vocabulary, the more she reacted to [ɑː] as being less typical than [a]. In the input distribution of /ɑ/ and /aː/ from Chapter 3, vowel sounds like [ɑː], with the vowel quality typically associated with the phoneme /ɑ/ and the vowel duration typically associated with the phoneme /aː/, were less frequent than vowel sounds like [a], with the vowel quality of /aː/ and the vowel duration of /ɑ/. The correlation between infants' vocabulary size and attention allocation to the atypical vowel sounds shows that by 15 months of age Dutch infants begin to develop fine-grained sensitivity to the auditory speech sound distribution in their language input.

6.3 Evaluating the role of computational models: Tools or theories?

The computational modeling in Chapter 5 was an in integral part of the program to investigate distributional learning in nature's distributional-learning experiment. The models were used to test whether infants could have acquired their representations of /ɑ/ and /aː/ (as tested in the speech perception experiments in Chapters 3 and 4) from the auditory frequency distribution of /ɑ/ and /aː/ in their input (as found in Chapter 3) through distributional learning. Distributional learning was simulated with Mixture-of-Gaussians (MoG) models, as well as with neural-network (NN) models that are embedded in a larger model for bidirectional phonetics and phonology (Boersma, 2007; Boersma et al., 2012). Both types of models induced the vowel contrast from the two-dimensional input distribution. The NN models were in this respect more robust than the MoG models. Both computational models associated the categories /ɑ/ and /aː/ with the respective vowel quality and duration of the categories. The models thus captured an important aspect of infants' early vowel representations: They are associated with multiple cues. Because of this similarity between the infants' and the models' representations, the question whether infants could have acquired their native language phoneme categories through distributional learning can be answered with "yes".

Of course, it is a qualified "yes", because the two models did not give exactly the same results. The extent to which each model accounts for infants' phoneme acquisition through distributional learning, can be evaluated in two ways.

The first evaluation compares the models' representations of /ɑ/ and /aː/ to what we conceptually think a model of distributional learning should acquire from these input data. The MoG model acquired two categories from the monomodally distributed duration distribution, which is not in line with the conceptual definition of distributional learning. This unexpected result was due to the Gaussian bias of the model. The NN model has no Gaussian bias and did not show this behavior. The second evaluation of the models is a comparison between the models' and infants' perception. Only the MoG models learned to treat [ɑː]-like sounds as infrequent and [a]-like sounds as ambiguous, a result that is in agreement with infants' perception (Chapter 4). The NN models predicted that infants could not acquire this difference between [ɑː]- and [a]-like sounds through distributional learning. This result in the NN model was the consequence of competition between the cues, which was absent in the MoG model. By evaluating the models against the input and perception data, I treat the models as theories that can be refuted or refined after an empirical test.

Earlier work that used computational modeling to test whether phoneme categories are learnable from the distributions in IDS through distributional learning primarily employed a Mixture-of-Gaussians (MoG) model (De Boer and Kuhl, 2003; Vallabha et al., 2007; Adriaans and Swingley, 2012). Although Vallabha et al. (2007) used a second, non-Gaussian model as well and also McMurray et al. (2009b) apply a Gaussian as well as a non-Gaussian model, the MoG approach to distributional learning is gaining popularity (see Chapter 5). In that line of work work, the MoG model is treated as a tool to help answer questions about the learnability of the input data or the dynamics of distributional learning. The MoG and NN models in Chapter 5 performed distributional learning on the same input data but differed in some outcomes. Therefore, neither can be regarded as a tool to model distributional learning without any assumptions.

In Chapter 3, I proposed to simply count the number of local maxima in an input distribution. According to the conceptual definition of distributional learning, the number of peaks should correspond to the number of categories that infants acquire. Although the results in Chapter 3 did not show two neat local maxima for /ɑ/ and /aː/, I believe that this method is important to explore in further research. In the first place, this method contributes to the comparison between the shape of the input distribution and the modeling results. In the second place, as long as theories and frameworks of infant language acquisition only adopt a conceptual understanding of distributional learning, their assumptions and predictions can best be tested with an assumption-free method.

The computational models in Chapter 5 were treated as specific theories about the distributional learning mechanism. The comparison between the MoG modeling and the NN modeling provided, among other things, a comparison between learning with and without a Gaussian bias. The comparison between models trained on onedimensional and two-dimensional distributions tested to what extent infants could acquire categories for individual dimensions (Boersma et al., 2003; Maye et al., 2008) or whether they should integrate all information that is available to them in order to acquire the speech sound categories (Pierrehumbert, 2003; Werker and Curtin, 2005). Such comparisons between learning scenarios require a definition of distributional learning that goes beyond a conceptual understanding of the mechanism. Only with a specific definition of distributional learning, theories and frameworks of infants' acquisition of speech sound perception can provide an explicit, computationally modeled, link between infants' input and perception. Such a level of specificity is available for distributional learning in the BiPhon model (Boersma et al., 2012), and for combined learning from speech sound distributions and the lexicon in Pierrehumbert (2001) and Feldman et al. (009b). Only with such explicit theories, the field can move to a formal understanding of distributional learning that is testable in nature's distributional-learning experiment.

6.4 Investigating infants' input: Against data reduction

Earlier studies of the phonetic properties of speech sounds in infants' input mainly focused on the enhanced contrast between category means in IDS as compared to adult-directed speech (ADS; e.g., Kuhl et al., 1997). The results in Chapter 2 on the realization of the corner vowels in Dutch IDS strongly suggest that the pronunciation of the corner vowels in IDS is language specific. Dutch mothers reduced their area of their vowel quadrilateral in IDS as compared to ADS. This reduction seemed to occur because the mothers fronted all their vowels in IDS as compared to ADS, but the back vowels more so than the front vowel. Such a shift of the vowel space can only be detected if the data are not reduced to the area of the vowel space and the analysis takes into account the actual average formant values of the vowels.

Apart from the fact that auditory contrasts between corner vowels are not universally enhanced, it is sub-optimal to measure auditory contrasts from only the category means and disregard the variance. Statistical techniques customarily evaluate the absolute differences between group means against a measure of the variation in the groups (*t*-test, Student, 1908; signal detection theory, Peterson et al., 1954; Tanner and Swets, 1954). Cristiá and Seidl (ress) have applied this insight to the measurements of auditory contrasts in IDS and have shown that conclusions about enhanced or reduced auditory contrast can change if variability is taken into account. This better measure of auditory contrast could not be used in Chapter 2, because of the low number of vowel tokens for some mothers in the adult-directed register.

Importantly, the higher mean F2 of the infant-directed corner vowels in Chapter 2 can serve as the starting point in explaining the observation that vowels are realized more variably in IDS than in ADS (Cristiá and Seidl, ress, and references therein). A higher F2 can be an acoustic consequence of smiling (e.g., Tartter and Braun, 1994) and a joyful smile is one of the three typical facial expressions in IDS (Stern, 1974; Chong et al., 2003). The other two typical infant-directed facial expressions are soothing protruded lips, and a surprised open mouth (Stern, 1974; Chong et al., 2003). Interestingly, these three facial expressions exaggerate the facial expressions that correspond to the articulations of the three corner vowels: 'smiled' [i], 'protruded' [u], and 'surprised' [a]. If mothers produce all corner vowels in IDS with these three infant-directed facial expressions, the realization of the corner vowels in IDS will be more variable than in ADS. I thus hypothesize that the affective speaking style in IDS is not only responsible for the shifts of the vowel category means, but also for the larger within-category variability. This hypothesis can only be investigated if the data are not reduced to one mean value per category.

As discussed in Chapter 1, a reduction of the input data to one mean per category is untenable in the investigation of the distributional-learning hypothesis. Distributional learning takes place over a range of auditory values and the shape of the frequency distribution is essential. The number of local maxima in a distribution can only be counted if the complete distribution is considered (Chapter 3). The results of the computational modeling of distributional learning would have been extremely uninteresting if only the means of each category had served as the input (Chapter 5).

To conclude, Chapter 2 shows that the vowel space should not be reduced to only a surface value if the effect of affect on the realization of phonemes is investigated. Chapters 3 and 5 illustrate that the categories should not be reduced to a mean if the distributional-learning hypothesis is tested. Therefore, I argue that phoneme categories in IDS are best studied with as little data reduction as possible.

6.5 Investigating infants' phoneme perception: Overt behavior and attention allocation

The reason for employing two tasks to test infants' perception of /ɑ/ and /aː/, a discrimination task in Chapter 3 and a categorization task in Chapter 4, was that both provide a different type of information about speech perception and have a different tradition in the research into infants' and adults' phoneme perception (see Chapter 1). Because the infants did not show anticipatory behavior in the categorization task, it is at this point impossible to evaluate the relation between discrimination and categorization with respect to Dutch infants' perception of /ɑ/ and /aː/.

In the discrimination task, infants showed evidence of discriminating between [ɑ] and [aː]. In the categorization task, infants failed to treat [ɑ] and [aː] differently. This could be the result of the difficulty of the task (cf. McMurray and Aslin, 2004). One advantage of the two-alternative categorization paradigm in Chapter 4 was that infants remained interested throughout the experiment and looked consistently to the trials. This allowed for the investigation of their attention allocation through pupil dilations. Not only did infants allocate their attention differently to atypical [ɑː] and [a] than to typical [ɑ] and [aː], the influence of context on the infants' attention allocation to [ɑː] and [a] mirrored the effect of context on adults' categorization of these atypical vowel sounds. In the discrimination task, a difference between the atypical vowel sounds [ɑː] and [a] was not found. Thus, a discrimination task can hide infants' sensitivity to the status of atypical speech sounds that attention allocation reveals. Because the procedures in Chapters 3 and 4 yielded null results and interpretable results for different vowel sounds, their combination proves once again how carefully behavioral null results must be interpreted, especially if we do no yet fully understand which processes are responsible for success in the task.

It has been argued that especially in infants, pupil dilations provide insight into a pre-conscious state of processing (Laeng et al., 2012). Infants' ability to differentiate between [ɑː] and [a] in their attention allocation in Chapter 4, but not in their behavior in Chapter 3, supports this. A combination of behavioral and pupillary analyses in future work will increase our understanding of the processes that underlie infants' behavior in speech perception tasks.

The studies to date that report infant pupil dilation results involved a clear point of potential surprise, either about the physical world (Jackson and Sirois, 2009; Sirois and Jackson, 2012) or the social world (Gredebäck and Melinder, 2010, 2011), or they were engaging with both audio and video (Lewkowicz and Hansen-Tift, 2012). The categorization task was relatively interesting, as shown by the low drop-out rate and by the infants tracking the boxes throughout the trials. Also, each trial could be split into parts during which arousal or a conflict in decision was expected to affect the infants' pupil dilations. The discrimination task, on the other hand, was relatively boring, as shown by the high drop-out rate. It is difficult to determine when during a long, monotonous trials infants should experience arousal from hearing an alternation between speech sounds. Moreover each infant was looking to the screen at different intervals during each trial. In order to take full advantage of the richness of pupil dilation in infant speech perception research, an extensive reanalysis of existing data sets is necessary. I expect that the paradigms that involve a clear point of potential surprise or are sufficiently engaging overall will allow for taking pupil dilation data into the equation.

6.6 Conclusion

Even though the parents of the participants in my studies were largely unaware of the learning task that their infant was accomplishing and spoke somewhat unclearly to their babies, they provided their children with distributions from which even computational models with basic distributional-learning mechanisms acquired the /ɑ/–/aː/ contrast. Not only infants' discrimination between typical examples of /ɑ/ and /aː/ but also aspects of infants' perception of atypical examples could be accounted for in terms of these models. The results in this dissertation therefore suggest that infants acquire the phonemes of their native language through distributional learning. This means that the phoneme categories emerge over the course of acquisition and are not innate. The results also show that an explanation in terms of distributional learning can only be maintained if infants can integrate multiple auditory cues during distributional learning. This finding requires further investigation of distributional learning in principle, with two-dimensional input distributions in an artificial-language learning experiment.

The results from all studies combined show that the research program that I called "nature's distributional-learning experiment", an integrated study of infants' input, infants' perception, and distributional-learning models to provide the explanatory link, is an essential contribution to testing the distributional-learning hypothesis of infants' phoneme acquisition in practice.