Hearing the Structure of Math: Use and Limits of Prosodic Disambiguation for Mathematical Stimuli

Michael Phelan

Department of Linguistics, The Ohio State University, USA

`phelan@ling.osu.edu`

Abstract

Listeners use the prosodic cues of an utterance to help determine its syntactic structure, but how does this process happen in the specialized domain of mathematics? Mathematical expressions can contain deeply embedded structures, and listeners encounter read mathematical expressions (RMEs) far less frequently than other potentially ambiguous utterances. How does experience with listening to math affect our ability to hear the structure of an RME via its prosody? Are there limits to the amount of structure we can pull out of the prosody of an utterance?

A perception experiment was conducted with subjects aged 7-59 to help answer these questions. Participants heard recordings of RMEs and attempted to determine which of two or more mathematical structures the reader intended. When subjects chose between two options for phrases like *nine times A minus two*, they chose the mathematical expression that had bracketing matching the prosody of the utterance. However, for more complex phrases like *the square root of sixteen over A plus twelve*, results were at chance. Age played a surprising role: subjects' performance increased dramatically from age 7 to 16, but adults' performance varied widely. This is attributed to variation in exposure to read mathematics.

Index Terms: speech perception, acquisition of prosody, prosody of mathematics

1. Introduction

If you have just heard someone say *two plus three squared*, should you be thinking of 11 or 25? The present study shows that the answer largely depends on the prosody of the utterance. Speakers use prosodic manipulations to communicate the intended grouping of words and phrases in their everyday English utterances, and listeners make use of these manipulations when determining which of several possible syntactic structures the speaker intended [1][2]. This paper extends these findings to the domain of read mathematical expressions (RMEs) and explores variations in the way listeners of different ages interpret certain prosodic phrasing. Finally, limitations on the use of prosodic phrasing for disambiguation are investigated with the use of complex RMEs that could correspond to more than two possible mathematical structures.

Read mathematical expressions provide and ideal test case for investigations of prosodic disambiguation for several reasons. First, the written form of a mathematical expression unambiguously shows the structural relationships between terms of the expression at a glance, while the utterance used to describe the mathematical expression can be ambiguous between two (1) or more (2) structures.

(1) Nine times A minus two: $(9 \cdot A) - 2$ $9 \cdot (A - 2)$

(2) The square root of sixteen over A plus twelve:

$$\frac{\sqrt{16}}{A+12} \qquad \frac{\sqrt{16}}{A}+12 \qquad \sqrt{\frac{16}{A}}+12$$

$$\sqrt{\frac{16}{A+12}} \qquad \sqrt{\frac{16}{A}+12}$$

Unlike everyday English stimuli, none of the mathematical structures corresponding to RMEs are inherently more or less plausible, and frequency effects are unlikely to bias listeners towards one selection or another.

Importantly for the purposes of this paper, readers in the same speech community have widely varying levels of experience with mathematical stimuli. Children who have long since attained adult-like levels of proficiency in using prosodic cues to disambiguate everyday English utterances may not yet have come across mathematical expressions like (2), and adults who do not work in education or math-heavy fields may not have dealt with such expressions in many years. On the other hand, older children and teenagers frequently encounter such expressions in the classroom, and thus may have an easier time disambiguating utterances of these expressions.

Finally, research to date on listeners' use of prosody for disambiguation has largely focused on whether a phrase should be interpreted with high or low attachment to previous parts of of the sentence. Comparatively little work has been done on prosodic disambiguation of utterances with many possible interpretations. The ease of generating such complex stimuli with mathematical utterances such as (2) can thus shed some light on the limitations of what listeners are able to disambiguate when forced to rely on prosodic cues alone.

2. Background and Assumptions

2.1. Previous research

Study of the prosodic disambiguation of syntactic ambiguities in non-mathematical speech goes back many years. Speakers have been shown to use lengthening and pausing to mark the intended structures of sentences with prepositional phrase and relative clause attachment ambiguities, complex noun phrase coordination ambiguities, and many others [3][4]. Listeners have likewise been shown to be sensitive to these

manipulations, at least when the interpretations differed in their syntactic bracketing [1][2].

A few previous studies investigated prosodic disambiguation of mathematical stimuli. Streeter [5] and Wagner [6][7] used simple mathematical stimuli like (3) and found that both speakers and listeners use the same prosodic disambiguation strategies as with non-mathematical stimuli.

(3)   a. $A + (E \cdot O)$
           b. $(A + E) \cdot O$

Only a few authors have used more complex mathematical stimuli, and all did so only in production. [8] and [9] both found longer pauses at locations likely to be marked with parentheses, though the only prosodic events coded by [8] were pauses over 300ms in length and [9] used a single speaker who consciously tried to prosodically disambiguate very complex stimuli.

In a production experiment with both simple (1) and complex (2) stimuli, [10] found that speakers produced RMEs using four distinct prosodic patterns. When intending structures like (3a), speakers most often used a strong prosodic break early in the utterance to group later terms together. When intending structures like (3b), speakers used a late prosodic break to group early terms together. Because the prosodic structure of these two sorts of utterances matched their syntactic structure, [10] referred to these utterances as having "cooperating prosody". In a few cases, speakers did the opposite, producing RMEs where the placement of breaks in prosodic structure grouped terms contrary to their grouping in syntactic structure, which [10] referred to as "conflicting prosody". Finally, many speakers produced evenly sized prosodic breaks throughout the utterance. Though the prosodic structure in these utterances did not necessarily group terms together in any particular way, different prosodically flat structures were used in different ways by speakers. When flat prosodic structures were used to indicate structures like (3b), the prosodic breaks were significantly more likely to be large, either all intonational phrase (IP) breaks or all intermediate (ip) phrase breaks. When flat prosody was used with structures like (3a), the breaks tended to be small ($\varnothing$, or prosodic word-level breaks). [10] referred to these prosodic patterns as "big flat" and "little flat". The current study uses stimuli from [10] to determine whether these four distinct prosodic patterns are interpreted by listeners the same way they are apparently intended by speakers, and whether age and the complexity of the mathematical expression affects listeners' ability to determine the structure the speaker intended.

2.2. Assumptions about prosodic processing

Accounts of the processing of prosodic breaks fall into two main camps: those assuming the Strict Layering Hypothesis of [11], and those like [12] that allow for recursion in the prosodic description of an utterance. Data from experiments like the one presented here can bear on this question, allowing investigators to test the limitations on what sorts of utterances can be reliably disambiguated via prosody. Both sorts of theories would agree that utterances of phrases like (1) should be easily disambiguated by listeners, while more complex phrases like (2) could quickly reach the limit of the number of

psychologically distinct levels of prosodic phrases in theories that assume the Strict Layer Hypothesis. The number of prosodic layers available to the speaker should set a hard limit on the degree to which multiply-embedded mathematical phrases could be reliably disambiguated. Under theories like [12] that are not subject to this restriction, speakers and listeners attempting to disambiguate utterances of phrases like (2) could make use of continuously varying relative differences in boundary strength. Drop-offs in perception accuracy for these more complex utterances would need to be explained by limitations on working memory capacity, and thus should be more gradual.

3. Experiment

To answer the questions raised above, a perception experiment was conducted with 29 subjects recruited from a local science museum. Subjects ranged in age from 7 to 59 (mean 25.2, sd 14.9) and participated in the experiment in groups of two to seven. None reported hearing problems, and all subjects whose data are presented here were native English speakers.

3.1. Stimuli

Three sets of mathematical stimuli were used. Two sets were selected from the recordings made for [10], in which naïve college students were asked to read twelve pairs of expressions like (4) and (5) with similar structures but different numbers.

(4)   a. $9 \cdot (A - 2)$
           b. $(10 \cdot A) - 2$

(5)   a. $\sqrt{\dfrac{16}{A + 12}}$

           b. $\sqrt{\dfrac{81}{A} + 72}$

Speakers in [10] were asked to avoid using terms like *the quantity, parentheses,* and *all of that,* and were told not to rearrange the terms of the expression, but were not told of the possible ambiguity in their speech. The two members of each pair (4a, b; 5a, b) were separated by at least four expressions with different structures, to prevent intentionally contrastive readings.

The first set of stimuli were chosen from pairs where one speaker produced cooperating prosody on both RMEs in a pair. The second set were chosen from pairs where a speaker produced cooperating prosody on one member of the pair and conflicting prosody on the other. A third set of stimuli were used in which one member of each pair was produced with IP breaks after each number or variable (called big flat, as in [10]) and the other member had only prosodic word level breaks throughout (little flat). Since such pairs were rarely produced by the same speaker in [10], these stimuli were recorded for this study by the experimenter.

There were 66 total mathematical items, 44 of which allowed only two structures as in (1), and 22 of which allowed either three or five possible structures, as in (2). All three sets

of stimuli were intermixed, such that the same expression was never repeated without at least five others expressions in between. The 44 "easy" items appeared first, followed by the 22 "hard" items.

A fourth set of stimuli consisted of eight non-mathematical English sentences like (6), which contained ambiguities due to prepositional phrase attachment, relative clause attachment, or complex NP conjunction.

(6)   Leslie photographed the manager with the iPad.

English stimuli were recorded by prosodically trained speakers unaffiliated with the study, who were told to disambiguate by putting an IP break either after the verb (for low attachment) or before the preposition (for high attachment) in PP attachment ambiguities and similarly for other ambiguity types. These eight non-mathematical English stimuli were given at the end to assess subjects' understanding of the general pattern of prosodic disambiguation in English.

3.2. Procedure

Subjects were given paper packets showing the possible mathematical structures that corresponded to each utterance. They were told they would hear people reading math problems and were to circle which problem they thought the subject meant, or, for the non-mathematical stimuli, which of two paraphrases they thought the subject intended. Stimuli were played back to groups of subjects via a laptop situated in front of the group, at a volume comfortable for all participants. There was a brief pause after each item to allow subjects to make their selections: six seconds following "easy" items, twelve seconds for "hard" items, and ten seconds for non-mathematical items.

4. Results

4.1. Mathematical results – Easy trials

Over all trials in which the subjects had a choice between exactly two mathematical structures, if the prosodic structure grouped two terms together, subjects were significantly more likely to pick the mathematical structure that grouped those terms together. Trials with cooperating prosody, where the grouping of terms in prosodic structure matched the grouping in syntactic structure, saw subjects selecting the intended mathematical expression on 67% of trials, significantly higher than chance (binomial test, $p < 0.001$). On trials in which the utterance was produced with conflicting prosody, where the prosody did not match the syntactic grouping the speaker intended to convey, subjects followed the prosody, choosing the mathematical expression that matched the prosody on 59% of trials (binomial, $p < 0.05$).

For the easy subset of problems with flat prosodic structures, listeners were significantly more likely to interpret "big flat" utterances containing IP breaks between each term as indicating left-branching syntactic structures (as in (3b)) than right-branching syntactic structures, doing so on over 80% of trials (binomial, $p < 0.001$). "Little flat" structures, however, were not consistently seen as marking one type of mathematical structure. Subjects selected left-branching and

right-branching mathematical structures at essentially the same rate (binomial, $p > 0.3$) on these trials.

4.2. Mathematical results – Hard trials

On trials where subjects were asked to pick between more than two answer choices, there was an significant difference in performance between problems that allowed for multiply-embedded constituents and those that did not. When subjects did not have to consider multiply-embedded structures, accuracy resembled that of easy problems, with choices following prosodic structure on over 60% of trials. However, when the utterance was consistent with multiply-embedded structures, as in (2) and (5), subjects were no better than chance at selecting the expression the reader had intended to convey (binomial, $p > 0.4$).

4.3. Non-mathematical trials

Overall accuracy on the eight non-mathematical items was reasonably high at 79%, significantly better than chance (binomial, $p < 0.001$). However 10 of the 28 subjects did score at or slightly below chance on these eight items, showing that understanding of the general English prosody-syntax correspondence rules is not universal. There was no significant correlation between age and accuracy on the non-mathematical items ($t = 1.717$, $p > 0.09$), though the two youngest subjects were the least successful on these questions.

In general, subjects who were most successful on the non-mathematical trials were also more successful on the mathematical trials ($t = 3.470$, $r = 0.56$, $p < 0.005$).

4.4. Interaction with age

There was an overall trend for accuracy in selecting the intended mathematical expression to improve with age ($t = 2.073$, $r = .37$, $p < 0.05$), though figure 1 shows that there is more going on than a simple trend to get better with age. Subjects improve dramatically from age 7 to 16, while adults are scattered all over the range. On non-mathematical trials adults did significantly better than children (64.4% vs 57.5%, $t = 2.234$, $p < 0.05$).
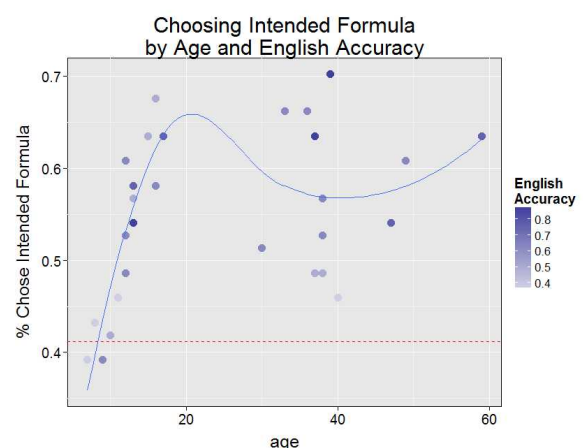

Figure 1. Effects of age and non-mathematical English trials accuracy in selecting intended form. Line of best fit is decidedly non-linear. Dashed line shows chance performance.

5. Discussion

In a listening experiment where they were asked to choose which of two or more possible mathematical expressions a speaker had intended to convey, subjects assumed that a larger prosodic break early in an utterance was meant to group later terms together, whereas a larger prosodic break later in an utterance was intended to group earlier terms together, as has been shown to be the case when listeners use prosody to disambiguate non-mathematical utterances.

Listeners correctly chose the intended mathematical form when there were only two options, or when there were more than two options but answers did not involve multiply-embedded phrasings. For trials where multiply-embedded structures were possible, listeners performed at chance. Under theories that follow the Strict Layer Hypothesis, this poor performance on complex problems could be attributed to listeners only have a small fixed number of psychologically distinct levels of prosodic phrasing. When the number of levels of embedding exceeds the number of prosodic phrasing levels available, disambiguation becomes difficult or impossible. Theories that reject the Strict Layer Hypothesis and allow recursion in prosodic phrasing would have to assume that this drop off in performance was due to higher memory load requirements for these complex mathematical expressions. The role of working memory in prosodic disambiguation could be tested by adding distractor tasks that increase memory load independent of the prosodic disambiguation task.

The striking effects of age evident in figure 1 suggests that familiarity with listening to or working with algebraic expressions like those used in the experiment has a strong impact on a listener's ability to use prosodic cues to disambiguate read mathematical expressions. The youngest subjects, just seven and eight years old, performed at or below chance. This was not surprising, given that these subjects were unlikely to have ever encountered expressions as complex as (4a,b), let alone (5a,b). [13] found that five to seven year olds could use prosodic breaks to choose between two different groupings in phrases like *pink and green and white,* but only when the pauses were quite pronounced (680ms), much longer than almost all pauses in stimuli here. Likewise unsurprising was the increase in performance for older children, who have had more experience with algebra and with prosodic disambiguation in general. What jumps out is the widely varying ability of adults in figure 1. While the results of the non-mathematical English trials clearly show that adults are able to use prosodic cues to disambiguate utterances, the presence of mathematical expressions seems to have greatly impacted adults' ability to apply the rules of prosody-syntax correspondence. Syntactically, the utterances used in the non-mathematical English trials are very similar or identical to those used in the mathematical trials, so differences between these two sets of results must be due to the fact that the stimuli were mathematical. It is unlikely that adults did not understand the structure of the mathematical expressions they were choosing between, since spacing, parentheses and other conventions of written mathematics indicate the grouping of terms quite clearly. It may be the case that adults, who generally do not have much exposure to algebraic expressions after graduating high school, are out of practice in applying known rules of prosody-syntax correspondence to the mathematical domain. There are few if any previous studies showing this sensitivity to a particular domain for prosodic disambiguation, and follow-up studies are underway with adult math teachers and college students majoring in math and engineering, who have much more exposure to mathematical expressions in their daily life.

6. References

[1]   Lehiste, I., Olive, J., & Streeter, L. Role of duration in disambiguating ambiguous sentences. Journal of the Acoustical Society of America. 60, 1199-1202. 1976.

[2]   Snedeker, J., & Casserly, E. Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. Language and Cognitive Processes, Vol. 25, 1234-1264. 2010.

[3]   Lehiste, I. Phonetic disambiguation of syntactic ambiguity. *Glossa* Vol. 7, 107-122. 1973.

[4]   Cooper & Paccia-Cooper. *Syntax and Speech*. Harvard University Press.1980.

[5]   Streeter, L. "Acoustic determinants of phrase boundary perception." *Journal of the Acoustical Society of America*, Vol. 64, 1582–1592. 1978.

[6]   Wagner, M. "Prosodic options or syntactic/semantic choices?"presented at the Workshop on Prosody and Meaning. Barcelona, Spain, 2009.

[7]   Wagner, M., & Crivellaro, S. "Relative Prosodic Boundary Strength and Prior Bias in Disambiguation." Speech Prosody, University of Chicago. 2010.

[8]   O'Malley, M., Kloker, D., & Dara-Abrams, B. "Recovering parentheses from spoken algebraic expressions." *IEEE Transactions on Audio and Electroacoustics,* Vol. 21 No. 3, 217-220, 1973.

[9]   Holm, B., Bailly, G., & Laborde, C. Performance structures of mathematical formulae. in International Congress of Phonetic Sciences. San Francisco, USA. 1999.

[10] Phelan, M. The Prosody of Algebra and the Algebra of Prosody: Prosodic Disambiguation of Read Mathematical Formulae. Proceedings of Speech Prosody 2012, Shanghai. 2012.

[11] Selkirk, E. O. *Phonology and Syntax: The relation between sound and structure.* Cambridge,MA: MIT Press. 1984.

[12] Wagner, M. Prosody and Recursion. Ph.D. Dissertation, Massachusetts Institute of Techonology. 2005.

[13] Beach,C., Katz, W., & Skowronski, A. Children's processing of prosodic cues for phrasal interpretation. *Journal of the Acoustical Society of America,* 99, 1148-1160. 1996.