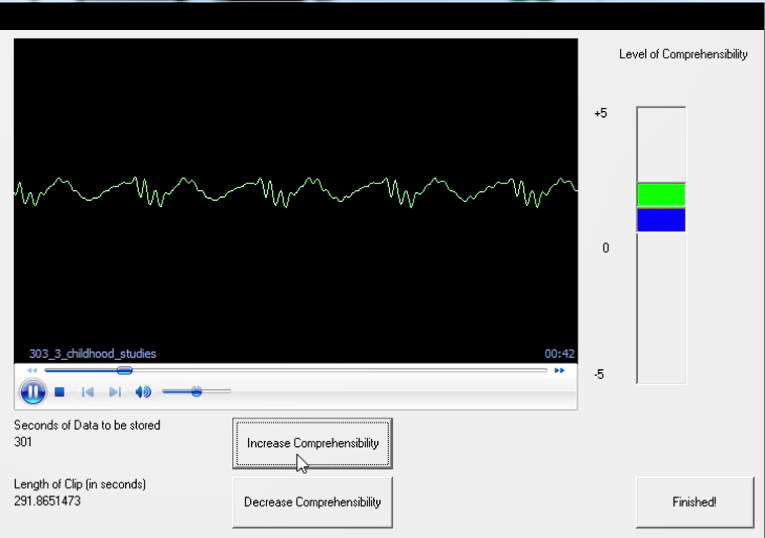
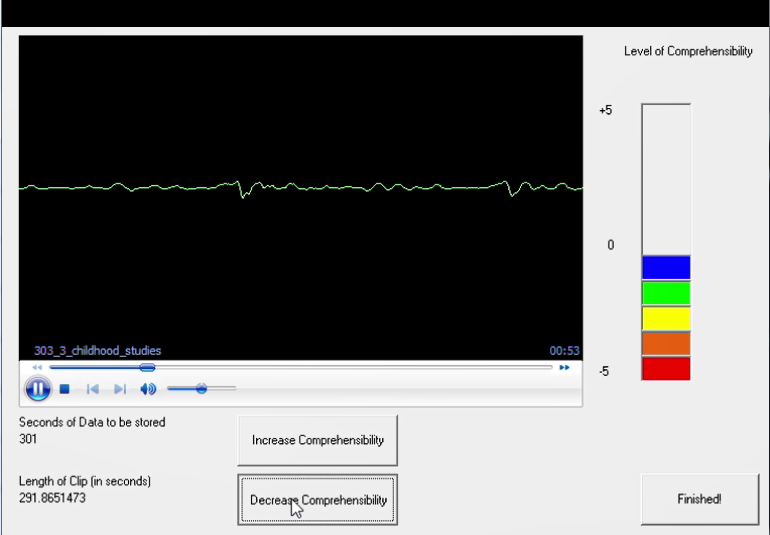
Appendix A

Screenshots of the Indiodynamic Software Interface





A listener clicking “Increase Comprehensibility” twice, which corresponded to a rating of +2 (top); a listener clicking “Decrease Comprehensibility” consecutively five times resulting in a rating of –5.

Appendix B

Debrief Questionnaire and Interview

Please rate your experience in today’s session by putting an **×** in the appropriate spot on the scale.

1. How well did you understand comprehensibility as a rating category?

I did not understand it at all

I understood this concept well

1. How comfortable did you feel when rating comprehensibility?

Very comfortable

Very uncomfortable

1. How difficult was the rating task for you?

Very difficult

Very easy

1. How confident are you in your ratings?

Not confident at all

Fully confident

1. Do you have any comments that you’d like to share about your rating?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Do you have experience with rating tasks like these? Have you ever done similar rating tasks? When?
2. What was the most difficult part of the rating task for you?
3. What was the easiest part?
4. Describe how you think your rating changed over time:

* as you were listening to each speaker
* from Speaker 1 to Speaker 3.

1. Which aspects of the speakers’ speech were most difficult to understand?
2. Which aspects were the easiest to understand?

Appendix C

Stimulated Recall Categories with Descriptors and Representative Examples

|  |  |  |
| --- | --- | --- |
| Category | Description | Examples |
| Discourse | Comments about the idea, the context, the narrative structure (sophisticated or simple), the details of the discourse, and the sophistication of the utterance content | 1. I got the idea more or less. I know what she wanted to say. 2. I think [I downgraded there] because the sequencing of the events didn’t make sense. |
| Lexis | Comments about lexical richness (rich and varied vocabulary or mostly simple words), lexical appropriateness (appropriate or inappropriate word choice), specific words or phrases | 1. She was trying to think of what to say, and instead of saying “World War II,” she used a strange word. 2. She used words that aren’t Spanish, but you can understand them: [school] subject is not “tópico.” |
| Grammar | Comments about syntax, grammatical complexity (elaborate or simple), and grammatical accuracy (especially gender-number agreement and subject-verb agreement) | 1. [She said] “Todas las días.” It’s “todos los días.” [In Spanish, *día*, ‘day,’ is a masculine noun]. 2. I think she said “fuimos,” it was the way she conjugated the verb. It wasn’t [correct]. |
| Fluency | Comments about speech rate (optimal, natural, too slow or too fast), pauses, and fluency in general | 1. I thought that the speech there was really fluent. I understood everything so [I upgraded her]. 2. Pauses are important. If you pause too much while you’re expressing an idea, I can’t follow you. |
| Incomprehensible | Comments about the listener’s incapacity to determine why comprehensibility diminished | 1. I couldn’t understand her. I’m not sure, I think she apologized, but I didn’t understand her well. 2. Of course [I downgraded her]. I didn’t understand anything there. I was totally lost. |
| Pronunciation | Comments about the production of individual sounds (frequency and type of error) and the speaker’s accentedness | 1. “Contruyimos” [listener imitates the pronunciation of the speaker]. I didn’t understand her pronunciation. It was the word. 2. She didn’t say it correctly, but I understood. The only error was the word “universidad” [‘university/college’], but the rest of the sentence was fine. |
| Rater experience | Comments about the listener’s identification with the speaker as an L2 learner or background knowledge the listener may have drawn upon to understand the speaker | 1. I understood her afterwards, that’s why I increased the rating. I understood because she elaborated on what she was saying and used a more sophisticated word. She tried to express herself more fully, like I do in my French courses. 2. I thought to myself, “does this affect comprehensibility?” And then I said to myself, “it depends who is listening.” It didn’t bother me since I knew she was looking for the word for “sand” but couldn’t remember it. |
| No comment | Comments from listeners who did not provide an explanation for clicking or who repeated what the interviewer said | 1. I upgraded her there. |
| Forgot reason | The listener does not remember the reason for clicking | 1. I don’t know why 2. Ah! I have no idea what I was thinking at the time. |
| Use of the software | Comments about issues with the use of the software | 1. I remember that my clicks were always a little bit delayed 2. Yeah, I think there was a slight delay when clicking. |

Appendix D

Frequency of Coded Comments (*k*) and Number of Raters Contributing Comments Through Stimulated Recalls as a Function of Rater Group (Dynamic, Semi-dynamic, Non-dynamic)

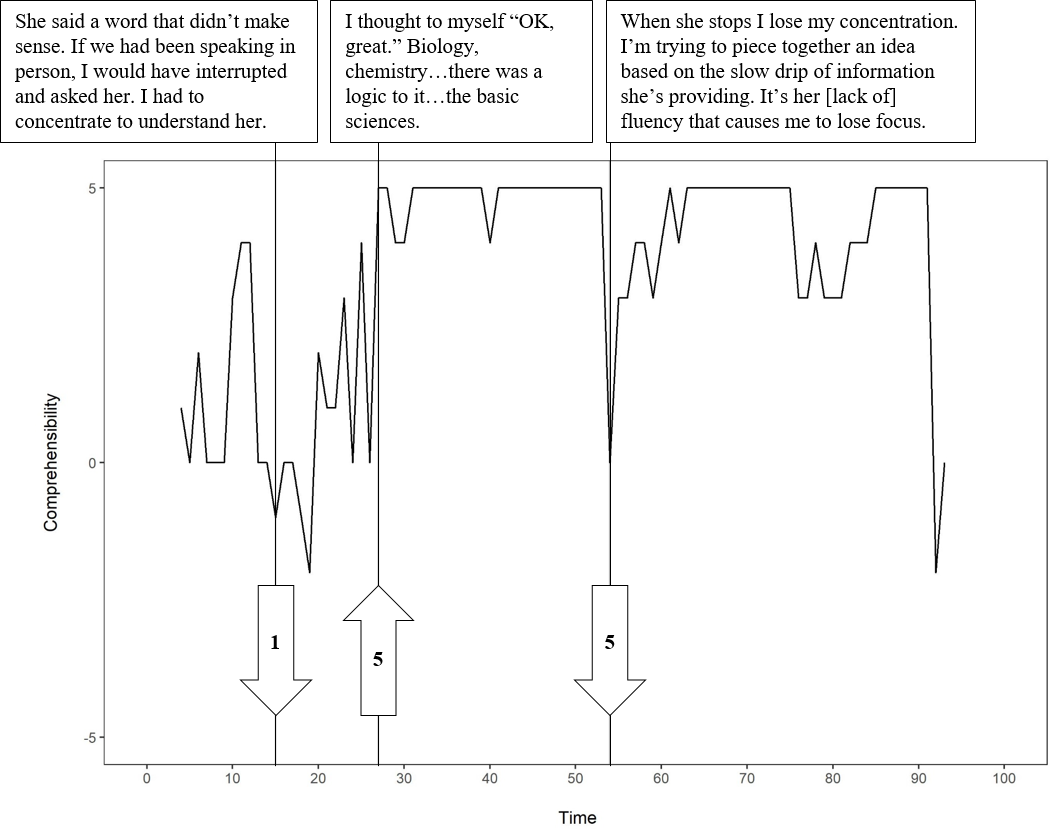
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Downgrade | | Upgrade | | No clicking | | Total | | |
| Coded category | *k* | Raters | *k* | Raters | *k* | Raters | *k* | % | Raters |
| Dynamic raters (*n* = 2) | | | | | | | | | |
| Discourse | 3 | 1 | 13 | 2 | 0 | 0 | 16 | 33 | 2 |
| Lexis | 5 | 2 | 0 | 0 | 0 | 0 | 5 | 10 | 2 |
| Grammar | 2 | 1 | 5 | 1 | 0 | 0 | 7 | 14 | 1 |
| Fluency | 1 | 1 | 12 | 2 | 0 | 0 | 13 | 27 | 2 |
| Incomprehensible | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 2 | 1 |
| Pronunciation | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 2 | 1 |
| Listener experience | 0 | 0 | 2 | 1 | 0 | 0 | 2 | 4 | 1 |
| No comment | 0 | 0 | 4 | 2 | 0 | 0 | 4 | 8 | 2 |
| Forgot reason | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Software use | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 12 |  | 37 |  | 0 |  | 49 | 100 |  |
| Semi-dynamic raters (*n* = 4) | | | | | | | | | |
| Discourse | 6 | 2 | 35 | 4 | 3 | 3 | 44 | 40 | 4 |
| Lexis | 9 | 3 | 7 | 3 | 0 | 0 | 16 | 15 | 3 |
| Grammar | 10 | 4 | 13 | 3 | 0 | 0 | 23 | 21 | 4 |
| Fluency | 1 | 1 | 3 | 2 | 0 | 0 | 4 | 4 | 2 |
| Incomprehensible | 6 | 2 | 0 | 0 | 0 | 0 | 6 | 5 | 2 |
| Pronunciation | 2 | 2 | 4 | 3 | 0 | 0 | 6 | 5 | 4 |
| Listener experience | 2 | 2 | 2 | 2 | 0 | 0 | 4 | 4 | 3 |
| No comment | 0 | 0 | 2 | 2 | 0 | 0 | 2 | 2 | 2 |
| Forgot reason | 1 | 1 | 1 | 1 | 0 | 0 | 2 | 2 | 2 |
| Software use | 2 | 2 | 1 | 1 | 0 | 0 | 3 | 3 | 2 |
| Total | 39 |  | 68 |  | 3 |  | 110 | 100 |  |
| Non-dynamic raters (*n* = 18) | | | | | | | | | |
| Discourse | 17 | 10 | 37 | 10 | 22 | 12 | 76 | 28 | 16 |
| Lexis | 32 | 13 | 6 | 5 | 14 | 9 | 52 | 19 | 18 |
| Grammar | 31 | 9 | 5 | 4 | 5 | 4 | 41 | 15 | 11 |
| Fluency | 11 | 7 | 5 | 3 | 5 | 5 | 21 | 8 | 11 |
| Incomprehensible | 20 | 12 | 0 | 0 | 2 | 2 | 22 | 8 | 12 |
| Pronunciation | 14 | 6 | 4 | 3 | 3 | 3 | 21 | 8 | 9 |
| Listener experience | 5 | 4 | 9 | 4 | 4 | 4 | 18 | 7 | 9 |
| No comment | 1 | 1 | 10 | 5 | 6 | 4 | 17 | 6 | 7 |
| Forgot reason | 3 | 2 | 1 | 1 | 1 | 1 | 5 | 2 | 3 |
| Software use | 1 | 1 | 0 | 0 | 1 | 1 | 2 | 1 | 2 |
| Total | 135 |  | 77 |  | 63 |  | 275 | 100 |  |

*Note*. No comment = instances when a rater provided no reason for rating. Forgot reason = instances when a rater could not recall or articulate a specific reason.

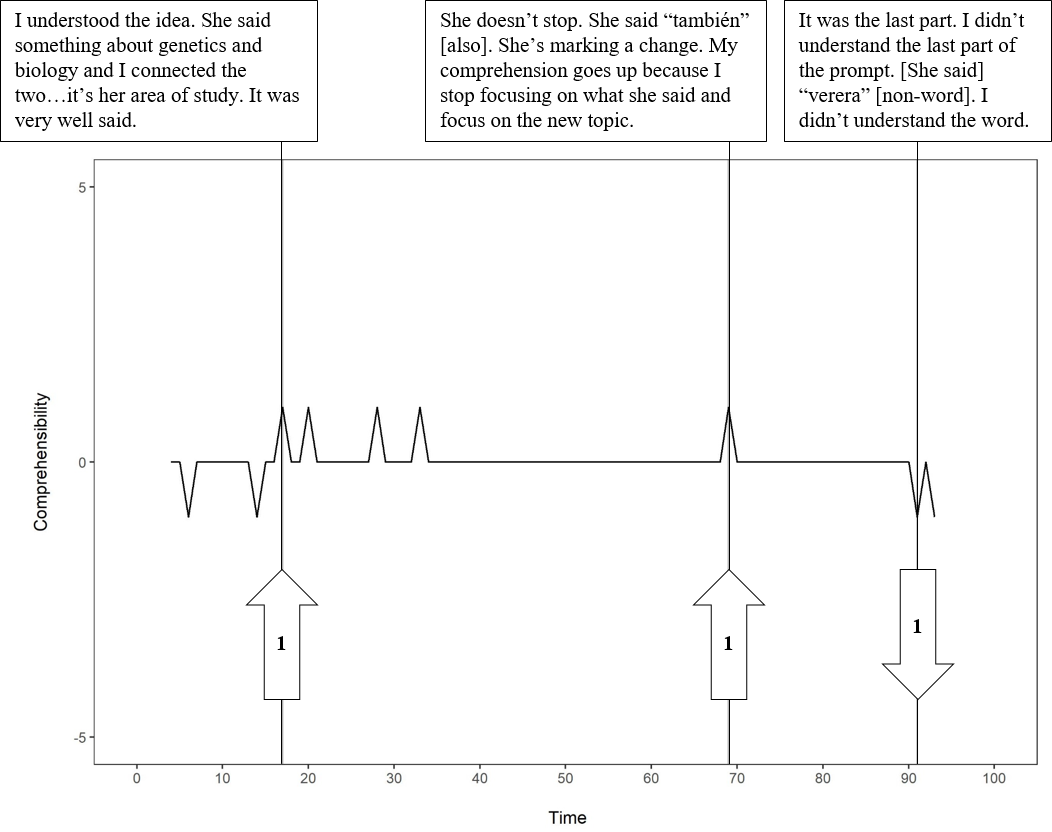
Appendix E

Contrasting Dynamic Versus Non-Dynamic Raters’ Responses

A particularly revealing contrast emerged when comparing the responses of dynamic and non-dynamic listeners for the same clip, which is illustrated here for the studies prompt by Speaker 2. As displayed in Figure D1 for a dynamic listener (Esteban), whose global rating for this clip was 8, Esteban downgraded the speaker slightly around the 15 second mark for using a word that he did not understand. This type of minor language use issue (e.g., using a word that does not make sense or would not be expected given the context) can be thought of as creating a temporary depression in comprehensibility, from which the listener is readily able to recover. In contrast, discourse structure and fluency features appear to be associated with more precipitous shifts in comprehensibility. In addition to their magnitude, such shifts may trigger an enduring state of high or low comprehensibility. For example, the speaker’s description of the courses she was taking, all of which were related to her major, resulted in a significant upgrade at about 27 seconds. Because the speaker was listing courses, both the content of the speech as well as its prosodic structure seemed to create a clear discourse structure for the listener, enhancing comprehensibility to a high level. Excluding minor perturbations, comprehensibility remained at this level until a fluency issue caused a downgrade to baseline at 55 seconds. In this case, the speaker had slowed down considerably, pausing often to contemplate what she wanted to describe as she was entering into more specific details about her coursework and path of study. Despite the sharp dip, comprehensibility was quickly restored, returning to its previous high state within the following 10 seconds.

*Figure D1*. Plot of stimulated recall comments and click behavior for a dynamic listener (Esteban).

Displayed in Figure D2, the response pattern for a non-dynamic listener (Juan Francisco), whose global rating for this clip was 6, illustrates some of the same themes, irrespective of differences in rating magnitude. When the speaker was beginning to list courses associated with her major around 18 seconds, Juan Francisco also upgraded comprehensibility. Much later, around 70 seconds, increased fluency combined with the use of *también* (“also”) triggered an upgrade. According to Juan Francisco, the transition signaled that he could stop trying to process what was previously said and instead allocate attention to upcoming information. Lastly, a mispronunciation resulted in a downgrade at the end of the clip. The speaker was describing her Spanish classes, noting that she found them interesting because they provided *algo de variedad* (“a bit of variety”). However, her pronunciation of *variedad* was unintelligible for Juan Francisco, who perceived something closer to *verera* (a nonce word). Taken together, the moment-to-moment assessment behaviors of the two listeners, provided in response to the same clip, appear to differ in the magnitude and extent of ratings and to some degree also in the stated reasons for these ratings, while overall following a similar temporal structure, with upgrades and downgrades occurring in comparable locations.

*Figure D2*. Plot of stimulated recall comments and click behavior for a non-dynamic listener (Juan Francisco).