2024. *Proc Ling Soc Amer* 9(1). 5691. https://doi.org/10.3765/plsa.v9i1.5691.

The time course of the rate of speaker transitions in conversation

David Edwards*

Abstract. Using over 500 hours of recorded conversations from the CallHome and CallFriend corpora (MacWhinney & Wagner 2010), this study analyzes conversations to measure gaps and overlaps in speaker transitions. It builds on previous work in (1) measuring speaker transitions by using custom software that measures single speaker segments based solely on acoustic evidence, without the need for a transcription or manual configuration, and (2) by breaking down the conversations and evaluating the transition rates minute by minute. Across the seven different languages studied, a pattern emerges in the rate of speaker transitions over the course of time. The number of transitions is highest in the first minute and gradually decreases before coming to a more consistent rate. This pattern exists in each of the languages studied, despite language-specific differences in turn-taking behavior. The cross-linguistic consistency of the decrease in the transition rate over the first several minutes of a conversation suggests that accommodation may be occurring as speakers negotiate the cadence of their interchanges with each other.

Keywords. turn-taking; corpus analysis; gaps; overlap

1. Introduction. In their ground-breaking work, Sacks, Schegloff, and Jefferson (1974) set out a list of observations of conversational turn-taking practices, a set of rules to explain those observations, and a transcription standard for capturing articulations and actions that are not captured in typical audio transcriptions. They ushered in the practice of Conversation Analysis (CA), which is now used by researchers in multiple fields to help provide detail and context for the actions that participants in a conversation make in response to each other. More recent research has attempted to quantify turn-taking phenomena, such as by measuring response times to polar questions (Stivers et al. 2009) or the overlap rate (ten Bosch, Oostdijk & Boves 2005).
This study expands on the methods of previous research used to quantify turn-taking with a process that can be used to analyze large audio corpora. Analyzing recorded telephone calls from two existing corpora, the current study also applies new metrics to provide additional lenses with which to view the distribution of transition gaps in a corpus. The analysis of transition rates by minute adds yet another means of making sense of the variations in interaction that are deployed by different speakers.

The by-minute analysis of transition rates provides a finding of this study that may stand out the most. Across the seven different languages studied, the number of transitions is highest in the first minute and gradually decreases before settling into a more consistent rate. This pattern exists in each of the seven languages studied, despite language-specific differences in turn-taking behavior, such as larger or smaller overlap rates or gap durations. Quantifying this pattern may help further the analysis of the intersubjective nature of conversation, studying the ways that speakers negotiate the cadence of their interchanges with each other as a conversation develops.

* I would like to thank Laurel Smith Stvan, Cynthia Kilpatrick, and Ivy Hauser for their assistance and guidance in the development of this research. Author: David Edwards, University of Texas at Arlington (david.edwards@mavs.uta.edu).

1.1. TIMING OF TURN-TAKING. Researchers have worked in recent years to quantify the measurements of turn-taking processes. Stivers et al. (2009) set out to measure the time between a turn and its response in 101 conversations in ten different languages. They extracted polar questions from the conversations and measured the time from the end of the question to the beginning of the answer. They found unimodal distributions with a mode between 0 and 200 ms for each language (with overlaps measured as negative gaps). The mean response times varied from a low of 7 ms to a high of 469 ms. Although this study has been crucial in establishing measurement and methodologies for measuring turn-taking gaps, the number of conversations in each language was not large. Thus, the exact values reported are perhaps less important than the general finding that the distributions of turns in various languages are similar and are centered on a point slightly above zero.

One differentiating factor in the timing of turn transitions is conversation modality, that is, whether a conversation is face-to-face or conducted via telephone. Ten Bosch, Oostdijk, and Boves (2005) showed that the overlap frequency in the Spoken Dutch Corpus, known in Dutch as Corpus Gesproken Nederlands, or CGN (Nederlandse Taalunie 2014), was lower in face-to-face conversation than in telephone conversations. More recently, sophisticated eye-tracking tools have allowed greater precision in measuring the gaze of conversation participants. Auer (2021) used these tools to evaluate gaze as a turn-allocation mechanism in multi-party conversations and concluded that gaze was an important and efficient tool for selecting the next speaker. For example, in what would be ambiguous situations for the selection of next speaker in the absence of gaze, the looked-at participant began the turn a majority of the time (Auer 2021: 88). This lack of visual cues in telephone communication makes that mode of communication less efficient than conversations held face-to-face, so a wider range of transition times should not be surprising.

Tian et al. (2023) compared Chinese conversations uploaded to a public website in three different modalities: face-to-face, online audio, and online video. Participants were familiar with each other, for the most part. The authors measured response times for question-response pairs, similar to Stivers et al. (2009). They found faster transitions (in the form of lower mean and median gaps) in the face-to-face conversations than in the online audio and video conversations. Their research provides evidence for the facilitation of in-person conversation compared not only to audio-only conversations but also to online video conversations.

1.2. LARGE-SCALE STUDIES OF SPEAKER TRANSITIONS IN UNSTRUCTURED CONVERSATION. With the availability of new software tools and greater computing power in recent years, it has become easier for researchers to analyze large corpora to identify turn-taking patterns. Since larger data sets are more likely to match the population in question, these automation tools allow us to have greater confidence that the measured phenomena represent attributes of the population as a whole.

The CGN study mentioned above (ten Bosch, Oostdijk & Boves 2005) studied 93 Dutch conversations. The recordings, which comprised over 15 hours of speech, included telephone and face-to-face conversations. The researchers used the word segmentation provided by the corpus, so they did not detail sound and silence parameters. Among other findings, they found that 52% of phone transitions were overlapped while only 44% of face-to-face transitions were.

Yuan, Liberman, and Cieri (2007) included conversations from the CallHome and Fisher corpora totaling over 500 hours of recordings. Using five minutes of each conversation, they used Praat, an acoustic analysis software tool, to identify sound and silence periods, without specifying what parameters they used. Among other findings, they found that Japanese had more

utterances overall and more short utterances (mostly backchannels) than the other languages.[1] They also found that there were significantly fewer overlaps in the Fisher corpus than in CallHome. The Fisher corpus had conversations between strangers, and CallHome had conversations between friends or family members. The researchers concluded that the familiarity of the participants had an effect on the overlap rate, with strangers overlapping each other less frequently.

Heldner and Edlund (2010) analyzed CGN (as did ten Bosch et al.), along with two map task corpora. They used a completely acoustic analysis method based on 10 ms frames. Their minimum sound duration was 90 ms, and their minimum silence was 180 ms. (Sounds less than the minimum sound duration were ignored, and within-speaker silences less than the minimum would cause the surrounding sounds to be treated as one.) They found that 40% of transitions were overlapped in CGN, including both telephone and face-to-face conversations. This is a lower percentage than what ten Bosch et al. (2005) found in either the telephone or face-to-face conversations.

Roberts, Torreira, and Levinson (2015) examined 348 conversations from the Switchboard corpus, which consists of conversations in American English among participants who did not previously know each other. The researchers made use of annotations from NXT-Switchboard but did not detail the acoustic parameters used to create the annotations. Excluding floor transfer offsets (FTOs, called *gap lengths* here), less than -2.2 s or greater than 2.2 s, they measured a mean FTO of 187 ms and a median of 168 ms.

Reece et al. (2023) introduced the CANDOR corpus, consisting of 850 hours of English conversations between previously unacquainted participants. Among other measurements, they found that 47.9% of transitions were overlapped, and the mean transition time was 200 ms with a mean of 80 ms. (The article does not describe the acoustic parameters they used to identify sound and silence segments.)

From these studies, we see that even when a large audio corpus is analyzed, the metrics are not always consistent, depending on the research questions that are being addressed. Most of these studies have reported gap lengths and overlap frequency, but only three of the five reported both. Sometimes gaps are treated separately from overlaps, and sometimes they are measured together. Although two of the studies make relative statements about transitions per minute, none of them report specific values for that potential metric. Evaluating the rate of transitions over time may give us insight into the dynamic nature of conversation as an interactive creation of interlocutors. It may also provide guidance on second-language teaching methods regarding differences in greetings, closings, or backchannels.

2. The current study.

2.1. AIMS. In addition, little research has been performed on how turn-taking processes change over the course of a conversation. One study (Yuan, Liberman & Cieri 2007) analyzed data starting with the third minute of each conversation they studied, but they did not offer a rationale for their selection of starting point. In addition to building on previous work by studying differences in turn-taking behavior among different languages, this study also seeks to measure how turn-taking behavior changes over the course of a conversation. Thus, this study is an exploratory effort seeking to determine what metrics may be most useful in quantifying turn-taking.

[1] The article does not include the exact percentages. The relevant figure indicates that for the turn-taking type of overlap (that is, not backchannels) the CallHome corpus has about a 15% overlap rate, and the Fisher corpus has about a 9% overlap rate.

2.2. DATA AND METHODOLOGY. The recordings for this study come from the portions of the CallHome and CallFriend corpora available via TalkBank (MacWhinney & Wagner 2010). They consist of phone conversations in seven languages: Arabic, English, French, German, Japanese, Mandarin, and Spanish. Participants in the calls in these corpora were allowed to call a person of their choosing anywhere in the world for up to 30 minutes. The relationship between the two participants was not recorded, but they are typically close friends or family. This study excludes conversations less than 12 minutes long or with excessive noise or other issues, leaving 1271 conversations, totaling over 544 hours of speech. See Table 1 for the details by language.

Language	Code	Conversations	Hours
Arabic	ara	180	82.00
German	deu	166	79.58
English	eng	227	106.66
French	fra	57	25.97
Japanese	jpn	168	71.20
Spanish	spa	254	101.77
Chinese (Mandarin)	zho	219	77.58
		1271	544.77

Table 1. Total duration of included recordings, by language.

A requisite task for this analysis is sorting out the sounds and silences. Because the recordings have the two sides of the conversation on separate audio channels, the determination of sound and silence states on each channel can be done separately, without risk of attributing a sound to the wrong channel. An automated process (Edwards 2023) was used to determine these sound and silence states on each channel using Praat (Boersma & Weenink 2022). The process used an intensity object based on a 200 Hz floor and a 4 ms frame step. With the intensity object, the sound and silence periods were determined based on an intensity threshold calculated for each file to yield a value representing 1.25% of the maximum that can be measured in a 16-bit audio file, or 52.2472 dB SPL. The minimum sound detection window was 30 ms,[2] and the minimum silence window was 200 ms. These settings were selected to filter out a portion of extraneous sounds (such as brief clicks generated by the telephone equipment or taps on a computer keyboard), to group together sounds separated by micropauses, and to capture some amount of paralinguistic vocalizations, such as audible inbreaths. With a 4 ms frame step, gaps measuring exactly 0 ms are infrequent (0.28% of all transitions), but they are still counted separately in order to avoid artificially inflating either the gap or overlap counts.

In addition, consecutive sounds without intervening sounds from the other channel were grouped together to identify single-speaker segments, in a similar fashion to previous research (ten Bosch, Oostdijk & Boves 2005; Weilhammer & Rabold 2003). This method differs from Heldner (2011) because it includes fully-overlapped segments as transitions. Basically, after the sound and silence segments are identified for each channel, the process groups sound segments together until a sound segment from the other channel occurs.

Extremely long gaps in these corpora are often the result of one speaker stepping away from the telephone. To prevent these instances from skewing the data, two adjustments have been

[2] With a 4 ms frame step, the effective minimum is 32 ms.

made: (1) any conversation with a gap longer than 30 s has been removed entirely, and (2) any gap between 3.5 s and 30 s has been reduced to 3.5 s. The latter change is intended to prevent these long gaps from artificially increasing the mean gap measurement while still taking the gap into account. Different studies of telephone conversations have used different maximums of gap lengths for their calculations. Examples include 2.2 s (Roberts, Torreira & Levinson 2015: 123), 2 s (Heldner & Edlund 2010: 559), and 3 s (Tian, Liu & Wang 2023: 6). The value of 3.5 used here is intended as an upper limit of what might be considered a normal turn-taking gap that does not involve stepping away from the conversation. On the other end of the gap length continuum, no maximum overlap has been set.

Because the final minute of a conversation may vary from the bulk of the conversation and because most conversations in these two corpora end abruptly at the recording limit (either 15 or 30 minutes), the by-minute analysis excludes the end of each conversation. The final minute is always excluded; even if the conversation reached the recording end, we don't know by the acoustic data alone whether they might have initiated their closing or pre-closing sequences (Schegloff & Sacks 1973). For conversations that end normally (before the recording limit), the final, partial minute is excluded along with the previous full minute. In this way, the by-minute analysis consistently excludes the end of every conversation.

2.3. TERMINOLOGY. Although *overlap* is used here similarly to other authors to indicate a period in which both channels are above the silence threshold, the nomenclature should not obscure the fact that overlap is in a continuous range with (positive) interspeaker gaps. Often, the only difference between a gap and an overlap is just a few milliseconds difference in the start time of the second speaker, and the difference between the two may come down simply to measurement precision. Thus, as part of a continuous variable, an overlap can also be considered a negative gap. Because the concept of negative value for a time measurement may not be intuitive and because *overlap* is a well-recognized term, this paper uses *overlap* in most cases in its text. However, graphs and tables label this measurement as *gap length* to indicate that it includes both positive and negative gaps. One alternative used by some researchers is *floor transfer offset* (FTO). However, in this paper, which is focused on acoustic measurements of conversations, FTO would not be accurate because in many cases, the transition is related to a backchannel contribution rather than an actual transfer of the floor from one speaker to the other.

In CA, a turn-constructional unit (TCU) arises out of the interactional exigencies at hand (Ford, Fox & Thompson 1996), but again, acoustic measurements have little insight as to what we might consider either a TCU or a turn. So although some analyses of acoustic corpora refer to turns, using a different term may avoid confusion. This paper will refer to a stretch of speech from one speaker uninterrupted by the other speaker as a single-speaker segment. Nevertheless, *turn-taking* is used to encompass actual conversational turns as well as backchannel contributions and paralinguistic vocalizations.

2.4. METRICS. Although previous studies have varied on the results they have published, five metrics together can describe the interaction between speakers: (1) between-speaker gap length, (2) overlaps per transition, (3) transitions per minute, (4) overlap duration per minute, and (5) between-speaker gap duration per minute. The latter four metrics exist just to provide additional ways of viewing the distribution of gap lengths (1). The number of overlaps per minute may be expected to increase in a conversation with a higher transition rate, so evaluating overlaps per transition (2) situates the overlap count within the context of the transition rate. The transition rate (3), then, becomes its own metric of how often each interlocutor speaks each minute. Although the latter two metrics can provide additional analytical value to the other three, they are

outside the scope of this paper. Then within these three metrics, evaluating how they change over the course of the conversation may provide additional insight into the behaviors involved in conversation.

3. Results.

3.1. GAP DISTRIBUTION BY LANGUAGE. Examining gaps found for each language, Figure 1 provides histograms by language for the selected parameters (see section 2.2). We do not see drastic differences in distributions of gap lengths among the languages. Each language has a modal value just below zero, with counts at the –1 s point (one second of overlap) being close to zero and somewhat higher at +1 s (one second of between-speaker silence). This difference in counts at equidistant points from zero illustrates the positive skew of the data (skew = .69).

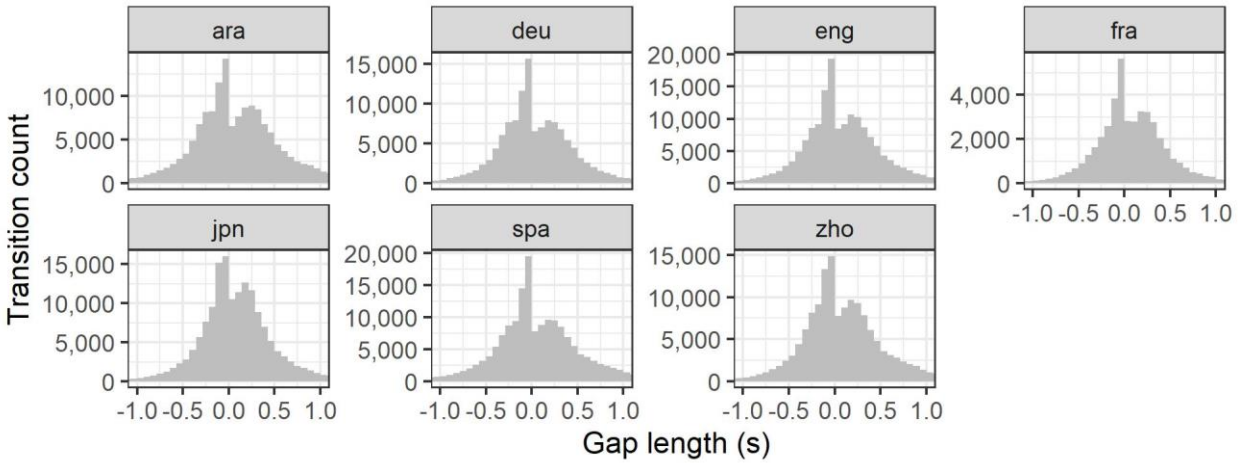

Figure 1. Gap length histograms by language, showing only -1 s to 1 s.

The difference between the sound duration threshold (30 ms) and silence duration threshold (200 ms) leads to an imbalance in the turn transitions, so that the largest frequency appears near –30 ms, and a secondary peak appears near 200 ms. Also, there is a slower decrease in slope on the positive side, showing the positive skew of each distribution. This positive skew can also be seen in Table 2, with the median value below the mean for each language. In a normal distribution, the median and mean values would be the same. Instead, the positive skew indicates that the conversations include a greater number of long gaps than long overlaps. This suggests that speakers in conversation tolerate long gaps more than they tolerate long overlaps.

Language	Mean (SD)	Median
ara	91(537)	44
deu	46(472)	-24
eng	98(502)	32
fra	81(445)	44
jpn	100(461)	60
spa	76(528)	0
zho	94(496)	36
Overall	85(499)	28

Table 2. Mean and median gap values in ms, by language

3.2. TRANSITION RATE BY LANGUAGE. Each speaker change is either a gap, an overlap, or a zero-gap transition. The number of such transitions per minute documents how often speaker alternation occurs, even if that alternation is just a short backchannel response. Figure 2 compares the transition rates for each of the seven languages.

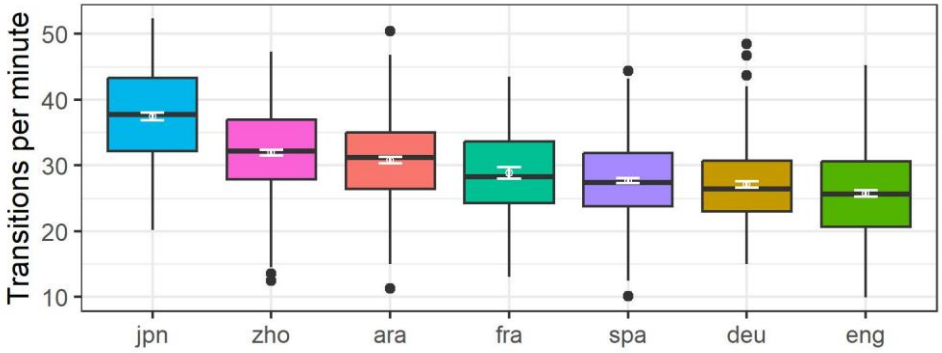

Figure 2. Transition rates, by language, with means and error bars in white

A one-way ANOVA indicates a significant difference in means of transition rates across languages ($F(6, 1265) = 62.38$, $p < .001$), and Japanese stands well apart from the others. (Scheffé post hoc tests demonstrates that Japanese differs from the other languages at an alpha level of 0.01.) In fact, the first quartile of transition rates for Japanese lies above the third quartile for German and English. That is, a phone call with a relatively low transition rate and relatively few backchannel responses in Japanese would still have more speaker changes than a relatively active English or German conversation. Table 3 lists the mean and median transition rates.

Language	Mean(SD)	Median
ara	30.84(6.76)	31.20
deu	27.09(6.42)	26.45
eng	25.75(6.94)	25.68
fra	28.88(6.63)	28.27
jpn	37.44(7.16)	37.79
spa	27.72(6.54)	27.37
zho	32.05(6.90)	32.27

Table 3. Transitions per minute, by language

3.3. OVERLAPS PER TRANSITION BY LANGUAGE. Because of the wide range in transition rates, comparing raw numbers of gaps and overlaps is less instructive than comparing the numbers of gaps or overlaps compared to the number of transitions. Since the number of zero-gap transitions is very small (0.28% overall), we can infer fairly closely the number of gaps per transition by looking at the number of overlaps per transition, and vice versa. Thus, the overlap frequency percentage, in conjunction with the transition rate, can serve as a proxy for both gap and overlap frequency. Figure 3 illustrates the overlap transition ratio for each language.

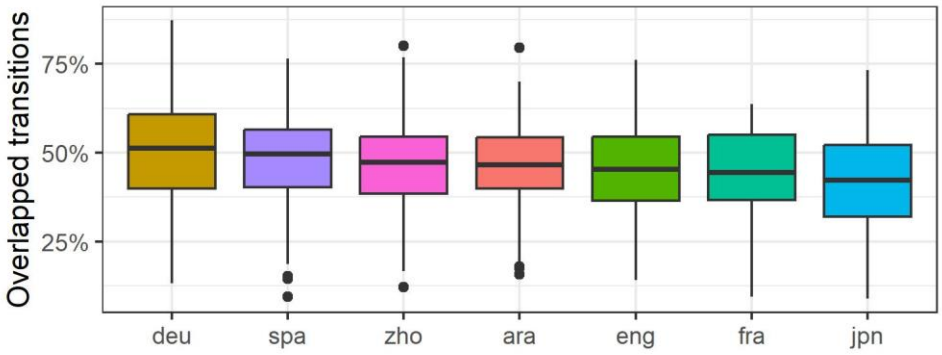

Figure 3. Overlaps per transition, by language

The 50% point (indicating near parity of overlap frequency with positive gap frequency) is within the interquartile range of each of the languages. Thus, in general, we can say that overlaps in these phone calls happen almost as frequently as gaps even when the transition rate varies widely. For example, Japanese has the lowest percentages of transitions overlapped despite having the highest transition rate. Table 4 shows the mean and median of each conversation's percentage of transitions that are overlapped. Despite the relative similarity that this metric exhibits compared to the transition rate, a Kruskal-Wallis test[3] shows significant differences between languages ($H = 39.33$, $p < .001$, df = 6).

Language	Mean	Median
ara	46.5%	46.6%
deu	50.6%	51.3%
eng	45.7%	45.4%
fra	45.3%	44.5%
jpn	41.9%	42.3%
spa	48.5%	49.6%
zho	46.5%	47.5%

Table 4. Transitions overlapped, by language

3.4. TRANSITION RATES BY MINUTE. Within this context of variation by language, Figure 4 shows the mean number of transitions per minute based on the beginning of the transition. There is a sharp drop-off after the first minute, and it continues to fall over the next three minutes, becoming more stable over the rest of the conversation.

[3] Kruskal-Wallis evaluates mean ranks and was used in place of ANOVA because the assumption of equal variances across groups was not met.

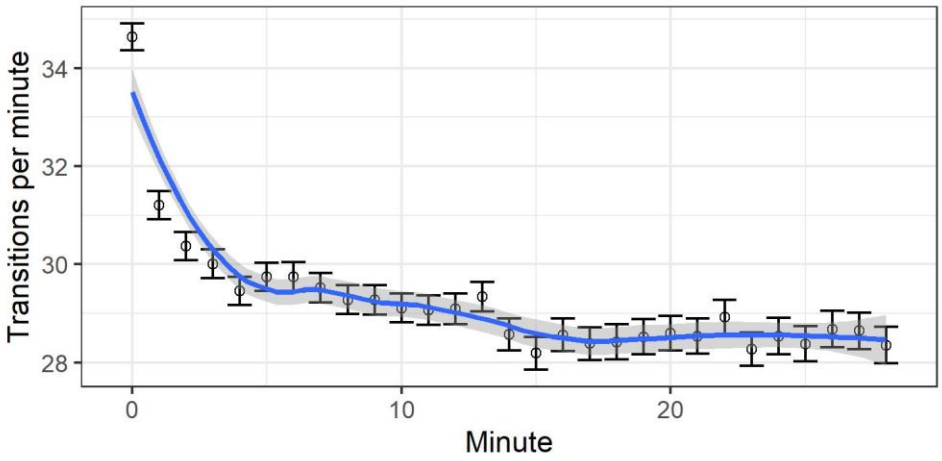

Figure 4. Transitions per minute, with error bars representing $\alpha = 0.05$

A similar pattern emerges in each of the seven languages, as seen in Figure 5, in which each loess line has an inflection point near the 4th or 5th minute. Although we see a consistently higher transition rate in Japanese, the loess line still shows a decline over the opening minutes before arriving at a steadier rate.

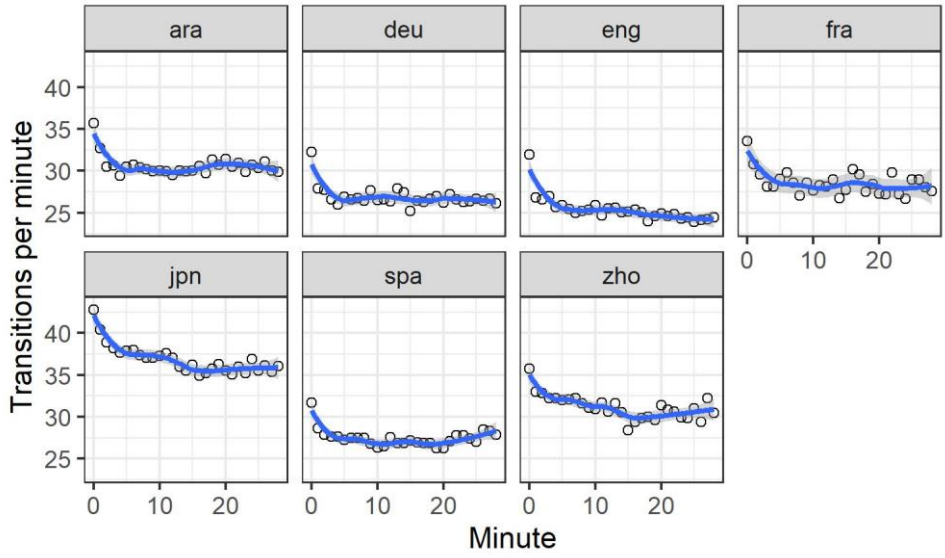

Figure 5. Transitions per minute, by language

In order to examine the effects of individual differences, a linear mixed effects model[4,5] was created with transition count (per minute) as the dependent variable. Because there is an inflection point in the loess lines in Figure 4 and Figure 5, a categorical variable was created for each

[4] This model was not created to test a hypothesis but instead to quantify the apparent effect in the figures above.

[5] Note that the assumption of equal variances was not met, but mixed effects models are more robust than ANOVA in handling this violation (Schielzeth et al. 2020).

of the first four minutes, with the remaining minutes being lumped together.[6] This last value was used as the reference value because it is the largest group. For the language variable, Arabic was used as the reference value simply because it comes first alphabetically. Random intercepts were generated for conversation number (treating each pair of speakers as a subject)[7]. Thus, the formula used was *trans_cnt ~ min4 + LangCd* + (1 | *conv*), where trans_cnt is the transition count (per minute), *min4* is the categorical variable described above, *LangCd* is the language code, and *conv* is the conversation ID. These variables are laid out in Table 5. The mixed effects model including time as a variable does not change the conversation-level pattern seen in **Table 3** above, in which Japanese has the highest transition rate and English the lowest. The model also indicates that the opening minute of these conversations has about 5 more transitions than the minutes from the fifth minute onward, as seen in the beta weight for the *min4b0* variable. (It should be noted, though, that this effect of a decreasing transition rate over the opening minutes of a conversation is seen in the aggregate and is not observed in all conversations.)

Variable	Description	β	SE	*t*
(Intercept)	est. transitions per minute for Arabic for 5th min. onward	30.305	0.505	60.002
LangCddeu	German compared to Arabic	-3.661	0.728	-5.029
LangCdeng	English compared to Arabic	-5.011	0.676	-7.418
LangCdfra	French compared to Arabic	-2.000	1.029	-1.943
LangCdjpn	Japanese compared to Arabic	6.614	0.728	9.090
LangCdspa	Spanish compared to Arabic	-3.163	0.661	-4.784
LangCdzho	Chinese compared to Arabic	1.162	0.684	1.700
min4b0	1st min. compared to 5th min. and after	5.364	0.212	25.252
min4b1	2nd min. compared to 5th min. and after	1.930	0.212	9.086
min4b2	3rd min. compared to 5th min. and after	1.115	0.212	5.251
min4b3	4th min. compared to 5th min. and after	0.729	0.212	3.434

Table 5. Fixed effects of linear mixed effects model estimating transitions per minute

In summary, the languages in these corpora show significant differences in transition rates and in the number of overlaps per transition. In particular, Japanese stands out with a higher number of transitions per minute than the other languages. In the by-minute analysis, each language exhibits a similar pattern of higher transition rate in the first minute that gradually falls over the next few minutes before settling in to a more-or-less consistent rate.

4. Discussion. The observation of a higher transition rate in Japanese corresponds with the findings of Yuan, Liberman, and Cieri (2007). This difference from the other languages can be attributed, at least in part, to expectations of *aizuchi*, a Japanese term that includes backchannel responses and non-verbal responses such as nodding (Kita & Ide 2007). Further, Japanese speakers are aware of the *aizuchi* of their interlocutors and may treat it as a topic worthy of discussion (Todd 2019). More research may be required to determine how robust the differences in transition rates are among the other languages. Still, the presence of a difference in backchannel and other response behaviors could be a valuable element to add to second language curricula.

[6] Although the loess lines in Figure 4 and Figure 5 may appear to continue to decrease after the fourth minute, the model was not improved by adding minute as a fixed effect. That is, no general effect of time was found.

[7] A random intercept for minute (representing different observations) did not improve the model and is therefore not included in the model presented here.

For the pattern of transitions at the minute level, multiple explanations are possible. One is that participants may have been hyperaware of the recording at the beginning of the conversation and became accustomed to it over time. One example is in (1), in which speaker A has been speaking for some time about the study that was undertaking the recording. Speaker B initiates a new topic, and seems to make use of clear speech, especially in the words *note* and *photo* in lines 04 and 05. It is possible that in this particluar case, with the recording having just been mentioned, the speaker was consciously trying to articulate distinctly.

(1) CallHome 4065: 84.6 – 92

```
01  B:          [ heh   heh   heh   heh  ]
02  A: yeah so [i'm guessing that's what] it's for and they need

03     a la[rge (  )]
04  B:     [aNyway  ] thank you very much for your for your noteʰ

05     [and your c]ard and your and your photʰo
06  A: [ohhhhh    ]
```

Since the protocol required the participants to consent to the recording prior to the beginning of the call, many (perhaps most) of the conversations explicitly refer to the recording in the opening minutes, as in (1). It is possible that this explicit, conscious awareness of the recording could have modified the turn-taking behavior in some way. However, the recording itself is typically only mentioned briefly before participants move on to other topics. The mention of it in the second minute of (1) is somewhat unusual. Further, it should be noted that participants would have presumably been taking part from their home or office, so the behavioral script invoked would be that of a typical telephone call, and the observer effect would presumably have been less pronounced than it would have been in a laboratory setting.

Another possibility is that opening sequences may explain the higher transition rate in the initial minutes of a conversation. Schegloff (1986: 117–118) identifies four sequences that are common to the openings of telephone conversations: summons/answer, identification, greetings, and "howareyou" sequences. In the CallHome and CallFriend corpora, the summons/answer sequence is slightly modified by the recording protocol. Then the remaining sequences proceed more or less normally, even if sometimes interrupted by comments about the recording process, as in 2). After the greeting sequence, they joke about the recording in lines 07 – 13 and then return to "howareyou" sequence in lines 14 – 18. Line 21 begins the first of what Schegloff calls "tellables," which speaker B begins elaborating on in line 29. From there, beyond the excerpt below, the transitions begin occurring at a slower pace. The first single-speaker segment of any length in (2) is from speaker B beginning in the middle of line 29. That segment begins not long after the 30-second mark, and at that point, 36 of the 46 transitions of that minute had already occurred.

(2) CallHome 1954: 0 – 36.8

```
01 A: [ aló]
          'Hello'
02 B: [mmar-]
      'mar(x)'
03    (0.6)
04 B: aló
      'Hello'
```

05 A: aló Eugenio
'Hello Eugenio'
06 B: () hola
'Hi'
07 A: hola me están grabando así que hablemos de cosas más coherentes
'Hi, they are recording me so we should talk about more coherent things. '
08 (0.7)
09 B: jajajaja
'hahahaha'
10 A: te pare[ce]
'What do you think? '
11 B: [°h] a es verdad es científico esto?
'Ah, is this scientific? '
12 A: sí es muy científico
'Yes, it's very scientific. '
13 B: jajaja bueno
'hahaha, good '

14 A: e có[mo ha esta]do? bien?
'And how have you been? Well? '
15 B: [°h] ((laughing))

16 (0.6)
17 B: bien (0.2) tú?
'Well. You? '
18 A: sí muy bien
'Yes, very well.'
19 B: ((x x))

20 A: [qué]
'What'
21 B: [choqué (0.4) choqué
'I crashed. I crashed. '

22 A: jajajajajajaja[jaja] (0.5)[jaja](0.6) [jajaja]
23 B: [jaja] °h [riese] ( ) riese aja
'Laugh Laugh'

24 A: ja[ja°h] bie[n por]qué
'haha good. Why? '
25 B: [jaja]ja [°h]

26 B: e-
27 (0.3)

28 A: [cómo.] [je je je]
'How? ' (laughing)
29 B: [°h] fui a [bajar] el río maipo en BALsa
'I went down the Maipo river in a raft – '

30 B: sí como rafting rafting como se dice así so[()]
'yes – how – rafting – how do you say rafting? – like () '

31 A: [ya] y?
'yeah, and? '

See Figure 6 for the transitions per minute of the entirety of conversation 1954, including the partial minute at the end of the conversation. Notice how the high transition rate drops drastically in the second minute.

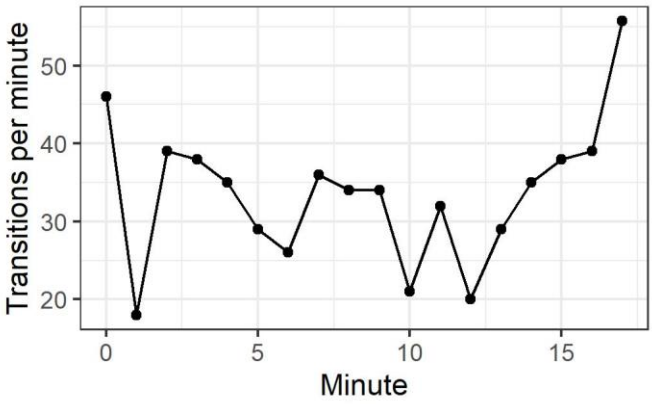

Figure 6. Transitions per minute for CallHome conversation 1954

What is happening in this conversation is that the narrative that speaker B begins at the end of (2) continues into the second minute with relatively few backchannels from speaker A. Then, as the conversation progresses, the transition rate moves between points between the two extremes set in the first two minutes. The final point on the figure represents a partial minute (just over 30 seconds). Although the patterns of the closing minutes are outside the scope of this paper, the conversations that include them tend to exhibit an increase in the transition rate as speakers carry out their pre-closing and closing sequences (Schegloff & Sacks 1973).

Opening sequences may explain the faster transition rate in the initial minute, but they would not seem likely to extend into the second or subsequent minutes. Since the second through fourth minutes also have a higher transition rate than the following minutes, a third possible explanation may also be playing a role: accommodation between the speakers regarding who is to speak and what topic they will discuss. Accommodation in speech rates has been documented (e.g. Cohen Priva, Edelist & Gleason 2017), but without a specific finding supporting accommodation of transition rates, this third possibility must remain speculative.

5. Conclusion. This study analyzed recorded conversations from two existing corpora using methods similar to previous research but with updated parameters. It added to that research a new metric: the analysis of transition rates by minute. Although the finding of a decreasing transition rate over the opening minutes of typical conversations might be entirely explained by the extended exchange of greetings (or perhaps, secondary greetings) that may occur in some conversations, it is also possible that additional accommodation work is being done, perhaps in both selection of speaker and of topic. To follow up on that, future research could look into whether people are trying out different communication strategies in the opening of a conversation before settling into one, or whether they continue to negotiate the rate throughout the conversation.

References

Auer, Peter. 2021. Turn-allocation and gaze: A multimodal revision of the "current-speaker-selects-next" rule of the turn-taking system of conversation analysis. *Discourse Studies* 23(2). 117–140. https://doi.org/10.1177/1461445620966922.

Boersma, Paul & David Weenink. 2022. Praat: Doing phonetics by computer. https://www.fon.hum.uva.nl/praat/.

Bosch, Louis ten, Nelleke Oostdijk & Lou Boves. 2005. On temporal aspects of turn taking in conversational dialogues. *Speech Communication* 47(1–2). 80–86. https://doi.org/10.1016/j.specom.2005.05.009.

Cohen Priva, Uriel, Lee Edelist & Emily Gleason. 2017. Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline. *The Journal of the Acoustical Society of America* 141(5). 2989–2996. https://doi.org/10.1121/1.4982199.

Edwards, David. 2023. Scripts to analyze audio files of conversations. https://github.com/utastudents/scripts-conversation-acoustic-analysis.

Ford, Cecilia E., Barbara A. Fox & Sandra A. Thompson. 1996. Practices in the construction of turns: The "TCU" revisited. *Pragmatics* 6(3). 427–454. https://doi.org/10.1075/prag.6.3.07for.

Heldner, Mattias. 2011. Detection thresholds for gaps, overlaps, and no-gap-no-overlaps. *The Journal of the Acoustical Society of America* 130(1). 508–513. https://doi.org/10.1121/1.3598457.

Heldner, Mattias & Jens Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics* 38(4). 555–568. https://doi.org/10.1016/j.wocn.2010.08.002.

Kita, Sotaro & Sachiko Ide. 2007. Nodding, aizuchi, and final particles in Japanese conversation: How conversation reflects the ideology of communication and social relationships. *Journal of Pragmatics* 39(7). 1242–1254. https://doi.org/10.1016/j.pragma.2007.02.009.

MacWhinney, Brian & Johannes Wagner. 2010. Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. *Gesprachsforschung* 11. 154–173. https://www.talkbank.org/.

Nederlandse Taalunie. 2014. Corpus Gesproken Nederlands – CGN [Corpus of Spoken Dutch]. http://hdl.handle.net/10032/tm-a2-k6.

Reece, Andrew, Gus Cooney, Peter Bull, Christine Chung, Bryn Dawson, Casey Fitzpatrick, Tamara Glazer, Dean Knox, Alex Liebscher & Sebastian Marin. 2023. The CANDOR corpus: Insights from a large multimodal dataset of naturalistic conversation. *Science Advances* 9(13). https://doi.org/10.1126/sciadv.adf3197.

Roberts, Seán G., Francisco Torreira & Stephen C. Levinson. 2015. The effects of processing and sequence organization on the timing of turn taking: a corpus study. *Frontiers in Psychology* 6. https://doi.org/10.3389/fpsyg.2015.00509.

Sacks, Harvey, Emanuel A Schegloff & Gail Jefferson. 1974. A simplest systematics for the organization of turn taking for conversation. *Language* 50(4). 696–735.

Schegloff, Emanuel A. 1986. The routine as achievement. *Human Studies* 9(2–3). 111–151. https://doi.org/10.1007/BF00148124.

Schegloff, Emanuel A. & Harvey Sacks. 1973. Opening up closings. *Semiotica* 8(4). https://doi.org/10.1515/semi.1973.8.4.289.

Schielzeth, Holger, Niels J. Dingemanse, Shinichi Nakagawa, David F. Westneat, Hassen Allegue, Céline Teplitsky, Denis Réale, Ned A. Dochtermann, László Zsolt Garamszegi & Yimen G. Araya-Ajoy. 2020. Robustness of linear mixed-effects models to violations of

distributional assumptions. (Ed.) Chris Sutherland. *Methods in Ecology and Evolution* 11(9). 1141–1152. https://doi.org/10.1111/2041-210X.13434.

Stivers, Tanya, N. J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, et al. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* 106(26). 10587–10592. https://doi.org/10.1073/pnas.0903616106.

Tian, Ying, Siyun Liu & Jianying Wang. 2023. A corpus study on the difference of turn-taking in online audio, online video, and face-to-face conversation. *Language and Speech* 00238309231176768. https://doi.org/10.1177/00238309231176768.

Todd, James Allen. 2019. "It has the ability to make the other person feel comfortable": L1 Japanese speakers' folk descriptions of aizuchi. *Lingua* 230. 102737. https://doi.org/10.1016/j.lingua.2019.102737.

Weilhammer, Karl & Susen Rabold. 2003. Durational aaspects in turn taking. *ICPhS* 15. 2145–2148. http://www.internationalphoneticassociation.org/icphs/icphs2003.

Yuan, Jiahong, Mark Liberman & Christopher Cieri. 2007. Towards an Integrated Understanding of Speech Overlaps in Conversation. In *ICPhS XVI*, 1337–1340. Saarbrücken.