Knowledge-Based Systems 71 (2014) 3–14

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Entrainment, dominance and alliance in supreme court hearings

Štefan Beňuš [a,b,*], Agustín Gravano [c,d], Rivka Levitan [e], Sarah Ita Levitan [e], Laura Willson [e], Julia Hirschberg [e]

[a] Constantine the Philosopher University, Nitra, Slovakia
[b] Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
[c] Departamento de Computación, FCEyN, Universidad de Buenos Aires, Argentina
[d] National Scientific and Technical Research Council (CONICET), Av Rivadavia 1917, C1033AAJ, Buenos Aires, Argentina
[e] Department of Computer Science, Columbia University, 450 Computer Science Building, 1214 Amsterdam Avenue, Mailcode: 0401, New York, NY 10027-7003, USA

ARTICLE INFO

Article history:
Available online 9 June 2014

Keywords:
Entrainment
Dominance
Coginfocom interfaces
Conversational fillers
Turn-taking
Linguistic style

ABSTRACT

A major goal of the Cognitive Infocommunication approach is to develop applications in which human and artificial cognitive systems are made to work more effectively. A critical step in this process is improving our understanding of human–human interaction so that it may be modeled more closely. Our work addresses this task by examining the role of *entrainment* – the propensity of conversational partners to behave like one another – in (1) the production of *conversational fillers* (*CFs*) and acoustic intensity; (2) patterns of turn-taking; and (3) Linguistic Style.

markers and how all of these relate to power relations, conflict, and voting behavior in a corpus of speech produced by justices and lawyers during oral arguments of the U.S. Supreme Court in the 2001 term. We examine several different measures of entrainment in justice–lawyer pairs to see whether or not they are related to justices' favorable or unfavorable votes for the lawyer's side. While two measures (a naive measure of similarity in CF rates and global similarity in CF phonetic realizations for the entire session) show no relationship, a third, which measures local entrainment in CFs in lawyer-justice pairs, does in fact identify a significant positive relationship between entrainment and justice votes. With respect to local entrainment in intensity, we found that lawyers do entrain more to justices than justices to lawyers, although there is no greater entrainment of female lawyers than of male lawyers. When we examine the relationship between entrainment in intensity and judicial voting, we find that, when justices voted for the petitioners, there is significant evidence of entrainment by both petitioners and respondents to justices. With respect to turn-taking behavior, we find that certain patterns of overlaps in turn exchanges between justices and lawyers are correlated with justices' voting behavior for four of the justices in our corpus. Finally, we find that there are lexical cues to divisiveness within the Court itself that can distinguish cases with close verdicts from cases with unanimous verdicts. We link these results to the possibility of building cognitive info-communication interfaces that exploit features of human–human entrainment for increasing effectiveness of human–machine interactions.

1. Introduction

One of the primary goals of the Cognitive Infocommunication approach [3] is to facilitate the development of "engineering applications in which artificial and/or natural cognitive systems are enabled to work together more effectively". One may approach this task by (1) improving understanding of the cognitive aspects of human–human interactions, (2) building formal models based on this understanding, and (3) implementing these models in human–machine systems to facilitate more natural and efficient interactions. In this paper, we report on a set of case studies aimed at the first step of this process in the area of speech ***entrainment***, the tendency of interlocutors to become similar to each other in terms of their acoustic and prosodic production (e.g. [41,6]). We examine how several types of such entrainment between conversational partners in the judicial domain relate to cognitive and social aspects of communication and information transfer.

A better understanding of entrainment is important for a number of applications of human–machine communication that rely upon Spoken Dialogue Systems (SDS). Research has shown that

[*] Corresponding author at: Constantine the Philosopher University, Štefánikova 67, 94974, Nitra, Slovakia. Tel.: +421 911 275318.
E-mail address: sb513@nyu.edu (Š. Beňuš).

http://dx.doi.org/10.1016/j.knosys.2014.05.020
0950-7051/© 2014 Elsevier B.V. All rights reserved.

not only do humans perceive conversational partners who entrain to their speaking style as more socially ***attractive*** and ***likeable*** [62,12,1], more ***competent*** (Street 1984), and more ***intimate*** [19], but interactions with partners who *unconsciously mimic* them are seen as more ***successful*** [22,60,35]. Different dimensions of entrainment have been shown to be reliable predictors of ***task success*** [64,55,49] as well. It has also been shown that humans may consciously decrease their similarity to others in order to increase their social distance to the interlocutor [34] or to show a negative attitude toward the interlocutor [13].

Not only do humans entrain to other humans, but studies have shown that they also entrain to computer systems. Nass and colleagues showed that human subjects perceive systems that entrain to them to be more likeable and interactions with those partners to be more successful [52]. A number of studies have shown that subjects do adapt to machines as well as to human conversational partners [5,25], so the ability to mimic this tendency found in human–human conversation would appear to be important for human–machine conversation as well, if SDS are to be as natural and effective as human partners.

It is well known that in spoken interactions between humans, information flows in multiple channels. Hence, information is construed as including not only the propositional semantic content of utterances but also other aspects of communicative functions such as, for example, Jacobson's referential, aesthetic, emotive, conative, phatic, and metalingual functions [44]. Cognitive Infocommunication approaches the notion of *channel* in a more abstract way as "a combination of sensory substitution and sensorimotor extension to convey structured information via any number of sensory modalities" ([26]: 261). We contribute to the notion of channel integration in future Cognitive infocommunication applications by examining how structured paralinguistic information contained in the acoustic channel of the spoken modality links to coordination and social relations between humans and might enhance the success of their interactions. The larger goal of this research is to replicate such behavior in human–computer interactions.

1.1. Entrainment

Coordination is a basic feature of human interactions. Sometimes, coordination in movements is explicitly intentional, or dictated by rules of social contact (for example in dancing) while other times it can be unintentional and facilitated by affordances of visual or aural modalities. For example, seeing somebody rock in a chair makes another person's rocking unintentionally entrained to this visual rhythmical movement [65]. Hence, visual information can couple with the movements of people involved in interaction and result in (unintentional) coordination.

Support for the coordination view of human–human spoken interaction comes from the literature on entrainment. For example, conversational partners tend to become more similar to each other as they speak. This phenomenon, known in the literature as entrainment, alignment, coordination, adaptation, unconscious mimicry, or 'the Chameleon Effect' (below we will use the term 'entrainment' exclusively), occurs along many acoustic, prosodic, syntactic and lexical dimensions – as well as in social behavior such as turn-taking – in both human–human interactions [15,25,64,77,54,78,18] and human–computer interactions [16,72,5]. Evidence of entrainment has been demonstrated in vowel spectra [1]; fundamental frequency [2]; pronunciation [40,59]; intensity [53]; voice quality [67]; lexical and syntactic choice [14,64]; frequency and duration of pauses [45]; speaking rate [39,5]; response latency ([21]); utterance length [50]; turn-taking behavior [47], jokes and laughter [63]. It has been found in many cultures: Hungarian [46], Frisian and Dutch [36,81], Hebrew [80], Taiwanese Mandarin [76], Japanese [79], Cantonese [32], and Thai [4].

1.2. Dominance

The notion of dominance is closely connected to entrainment between conversationalists. Thus, the amount of entrainment among interlocutors represents a potential window into the dynamics of the power relationship. The observed differences between speakers in their degree of entrainment to different individuals may indicate an asymmetrical power relationship. More specifically, if one interlocutor adjusts his or her behavior to that of a conversational partner more than the partner does, the former is likely to be perceived as playing a less dominant role than the latter. The understanding of asymmetrical distribution of power, and related aspects of dominance and status and how they are signaled through speech, has great potential for facilitating the quality of interactions between natural and artificial cognitive systems, since they represent a natural component of human–human interactions.

This view of an asymmetrical power relationship dynamically created or maintained through communicative interaction is closely related to the *dyadic power theory* of Dunbar and colleagues [29,20,30]. In this theory, dominance is seen as a combination of personal and contextual characteristics. The personal characteristics are the constant features of an individual which can be considered as personal traits that are independent of the situation with which the individual is faced. The contextual characteristics include the dominance or submissiveness of the individual's partner in the interaction. Here, we follow Poggi and D'Errico [61] and view dominance as a dynamic multidimensional communicative act by means of which one's interlocutor exerts power or influence over one or more conversational partners by displaying linguistic signals of dominance.

In Poggi & D'Errico's view, dominance is dynamic: it evolves over time. Interlocutors may begin an interaction with roughly equal power positions and finish with very different ones. Alternatively, the power relationship may be similar at the beginning and end of the conversation but may diverge in various dimensions during the conversation. In this sense, dominance is constantly being negotiated. Dominance is also multidimensional in that one conversational partner may be more dominant than their interlocutor in one dimension or aspect of the conversation while the roles might be reversed in other dimensions or aspects. Dominance is also relational, and is not assessed in absolute terms; rather the dominance of an interlocutor is only defined in relation to the dominance of their conversational partners.

Here we construe the ability or intention to influence a conversational partner as observable in the use of spoken language during the conversation. We hypothesize, following Giles et al. [33], that the degree of entrainment in speech will be asymmetric, and the less dominant speaker will entrain more to the more dominant speaker than vice versa. Assuming that social status can be principally linked to power and dominance, support for this hypothesis comes from studies of non-verbal behavior by Gregory and Webster [38] which showed that lower status partners entrained their voices to higher status partners.

1.3. Entrainment, dominance, and the judicial domain

Most studies investigating the relationship between entrainment and dominance have analyzed corpora collected in the laboratory or in situations where the 'stakes' were relatively low – that is, neither party was heavily invested in the outcome of the conversation. In such situations, the relationship between entrainment and dominance is hypothesized to form a social glue, indirectly facilitating successful outcomes in the low-stake tasks at hand (e.g. collaborative dyadic dialogs). Knowledge gained from such studies is useful in collaborative human–machine interactions geared toward providing information or accomplishing service tasks such as a reservation service. However, applications based on cognitive info-communication are likely to be utilized also in domains with high stakes such as crisis management (terrorist attacks, natural disasters, or fires), political recruiting calls, or even advertising campaigns where the product is expensive.

In this paper, we explore the domain of U.S. Supreme Court oral arguments, in which the stakes are high for participants. Although justices may form their opinions based on submitted briefs, research, and contributions of *amici* (lawyers speaking for one side or the other but not officially petitioners or respondents), effective oral argumentation may still win or lose a case for lawyers on both sides (*petitioners* and *respondents*). Hence, justices may come to the oral argument session undecided or even change their opinion based on the oral argument and a persuasive lawyer can make a difference in a close case. Additionally, both lawyers and justices have professional and very public faces that they strive to maintain. This is presumably more important for justices, given their high social status.

Previous research has found that justices and lawyers entrain in terms of their linguistic style, defined in terms of their patterns of function word use, and that this coordination matches the power relationship between justices and lawyers: Lawyers match their style to justices more than justices do to lawyers in their oral arguments before the Court, mirroring the fact that justices are more powerful. Moreover, lawyers were found to coordinate more with judges who end up voting against them than with more favorable judges [27]. In this paper we examine other forms of communicative behavior to identify additional evidence of entrainment and their relationship to power relations and Court decisions. We explore the extent to which lawyers entrain to justices and how this relates to the intrinsic power relationship between the two groups and its consequences in judicial decisions. Although dominance and entrainment have an inherent dynamic nature, as discussed in the previous section, our operationalization of dominance in this paper relies on a more static notion of dominance stemming from the power asymmetry between justices and lawyers. This allows us to extend the line of research in Danescu-Niculescu-Mizil et al. [27] and present comparable observation related to speech-based rather than text-based features of dialogs. After establishing this groundwork, the investigation of dynamic changes during the course of interactions is planned for future research.

The rest of the paper is structured as follows. Section 2 describes the corpus of the U.S. Supreme Court oral arguments and summarizes our measurements. Section 3 describes four experiments exploring the relationship between communicative behavior and social aspects of dominance and alliance: entrainment in the choice and segmental realization of conversational fillers and its effect on voting, entrainment in acoustic intensity and its relation to social status, justices' and lawyers' turn-taking behaviors and links to power relationships, and the employment of linguistic style markers for alliance and its presence or absence within the group of justices. Due to differences between the studies, methodological aspects of labeling, features, and their extractions are discussed separately with each experiment. Section 4 discusses the results, linking them to the research agenda of cognitive info-communications, and presents our conclusions.

2. Method

2.1. Corpus

In this paper we analyze data from the recordings of the 2001 term of the U.S. Supreme Court oral arguments consisting of 76 arguments, each of which takes approximately one hour. According to Court rules, each set of oral arguments lasts no more than an hour, with a 30-min time limit for both the petitioner and the respondent. This time limit is strictly enforced by the Chief justice and thus has a considerable effect on turn-taking strategies: justices routinely interrupt lawyers with questions, and lawyers also interrupt justices to salvage as much of their limited time allotment as possible. In the 2001 term, William Rehnquist (REHN) is the Chief Justice, and the remaining justices are Stephen Breyer (BREY), Ruth Bader Ginsburg (GINS), Anthony Kennedy (KENN), Sandra Day O'Connor (CONN), Antonin Scalia (SCAL), David Souter (SOUT), John Paul Stevens (STEV), and Clarence Thomas (THOM). Because multiple lawyers can represent plaintiffs and defendants, 198 lawyers appear in this term, and since lawyers may appear in multiple cases, the corpus includes speech from 150 unique lawyers.

The 2001 term recordings form part of the Scotus Corpus of U.S. Supreme Court sessions, which have been transcribed and time-aligned by the Oyez project [42]. We also use the U.S. Supreme Court Database [43] to document individual cases and justices' votes for or against plaintiffs and respondents. The 2001 term sessions were manually transcribed (including disfluencies and non-lexical speech) by professional transcribers and also manually sentence-aligned by the Linguistic Data Consortium, which also identified speaker turns. These turns are defined as consecutive words from a single speaker, ignoring any silent pauses; when two (or more) people speak at once, the overlap is considered its own turn with two (or more) owners. There are 24,910 turn exchanges in the 76 oral arguments in this corpus. 17,729 (71%) contain no overlap between adjacent speaker turns; in 7177 exchanges (29%) there is an overlap between two speakers; and in only four, three speakers speak at once. For reference, the proportion of overlapping exchanges is slightly lower than the one found in the Columbia Games Corpus, a collection of spontaneous collaborative task-oriented dialogs elicited from native speakers of Standard American English [37], in which roughly 33% of all turn exchanges contain some overlapping speech. More detailed analyses of overlap types is given in Section 3.3 below.

2.2. Measurements

Our studies employ a variety of measurements to examine relationships between the words lawyers and justices use and the way they produce them and the dominance relations, justice voting behavior, and degree of contentiousness we can observe in the corpus. Our examination of entrainment patterns makes use of three measurements of entrainment which we have defined in previous work. The first measure follows the approach of Nenkova et al. [55] and allows us to examine the similarity in the frequencies of usage of lexical items such as conversational fillers. Using the transcripts, we obtained the frequencies of lexical items d and total word counts separately for each oral argument, and calculated entrainment in filler type frequency as in ENTR1 (1).

$$ENTR1(d) = -\left|\frac{count_{S1}(d)}{ALL_{S1}} - \frac{count_{S2}(d)}{ALL_{S2}}\right| \quad (1)$$

In this definition, d represents a particular lexical item or class of item; $S1$ and $S2$ represent a pair of speakers; $count_S(d)$ is the number of times speaker S used d, and ALL_S is the number of all words uttered by S. Entrainment is thus the difference between the frequency of d in $S1$ and $S2$'s conversation.

Two other measures of entrainment make use of the Euclidean distance between the means of some variable d to measure similarity between $S1$ and $S2$ over a conversation. In ENTR2, we measure the Euclidean distance between values of d for speaker pairs $S1$ and $S2$ vs. $S1$ and $S3$. If this distance is smaller between $S1$ and $S2$ than it is between $S1$ and $S3$, we say that $S1$ is more similar to $S2$ than to $S3$ in this dimension d. This is a global measure of entrainment, over an entire session in our corpus. In the third entrainment metric we employ, ENTR3, we assess the similarity of $S1$ and $S2$ along some dimension d more locally. Our corpus of oral arguments frequently includes multiple short dialogs between a lawyer and a justice consisting of multiple turns. ENTR2 would not detect whether a lawyer adjusts his or her speech to a particular justice in such local dialogs. We explore this type of entrainment by calculating the Euclidean distance between speakers' productions for only adjacent productions of d for each pair of speakers. We term this form of entrainment ENTR3. We make use of another measure of local entrainment previously proposed in [47]: For each exchange between $S1$ and $S2$ we compute *Inter-Pausal Units* (IPUs). These are defined as pause-free segments of speech from a single speaker which is surrounded by silence longer than 50 ms. For each turn between $S1$ and $S2$ in the corpus, we compare the first IPU of $S1$'s turn with the last IPU of $S2$'s preceding turn, taking the negated absolute value of the difference between the two segments' values along some dimension. This measure describes the extent to which one speaker matches his interlocutor when beginning a new turn. We term this metric ENTR4.

In the experiments described below we also make use of standard statistical techniques (frequency counts, histograms, contingency tables, t-tests, logistic regression) to describe relative frequencies of lawyers' and justices' productions, to plot characteristics of some productions in two-dimensional space, and to identify possible relationships with dominance patterns and voting behaviors.

3. Experiments

3.1. Choice and segmental realization of conversational fillers

Despite the sometimes claimed absence of semantic content, *conversational fillers* (CFs) such as *uh* or *um* fulfill an array of discourse and pragmatic functions. For example, fillers heighten the attention of the listener and consequently facilitate the retention of information in memory [73]. Moreover, interlocutors have been found to entrain on them [7], and their timing has been found to be important in turn-taking coordination [8]. Finally, fillers also facilitate both production and perception of linguistic material because they allow speakers to plan their intended message and listeners to prepare to perceive important content. Listeners have been shown to be highly sensitive to the occurrence and timing of CFs in speech; for example, they inform the listener about meta-cognitive states of the speaker [17] and participate in structuring discourse (e.g. [74]. Hence, conversational fillers play an important role in inter-personal communication.

In addition to these functions, CFs might provide insights into the coordination and entrainment between interlocutors. Pentland [58] suggests that the communication system between humans has developed primarily through coordinating behaviors among people. The social aspect of this behavior is facilitated by so-called honest signals that enable this coordination through the utilization of redundancy of language in terms of temporal patterns, intensity, and fundamental frequency, since these features can be deployed to convey both linguistic and paralinguistic information. In other words, the acoustic signal of spoken communication provides multiple affordances for inter-personal coordination. Pentland further assumes that this cognitive coordination system operates orthogonally from the semantic content of our utterances, since honest signals seem to pre-date language evolutionarily. If we follow Pentland in assuming that coordination among conversational partners is guided by honest signals, the low semantic content of CFs should provide an ideal site for latching of these honest signals.

Additionally, previous research has shown that the quality of vowels provides an acoustic affordance for entrainment (e.g. [28,59], and thus the acoustic elements in fillers, lacking functional importance, represent a good testing ground for the functioning of these communicative social signals. Finally, fillers occur frequently in spontaneous speech, they are typically delimited by silent intervals, and their acoustic and prosodic characteristics are deeply redundant.

Here we examine the question of whether the segmental realization of CFs uttered by Supreme Court justices and lawyers presenting oral arguments before the Court vary significantly with the votes of each justice. That is, we ask if justices tend to vote in favor of the lawyers who entrain to them in terms of their production of CFs.

3.1.1. Conversational fillers in the Scotus corpus

Our corpus includes a number of filler markers. In this section, we focus on four of the most common markers: *ah*, *uh*, *eh*, and *um*, representing 99.9% of all filler tokens and 2.29% of total words in the corpus. Their relative frequencies are shown in the third and fourth columns of Table 1. This 2.29% rate of CFs to total words is comparable to the rates of CFs in other corpora (e.g. [69]).

We note from Table 1 that CFs in our corpus are more often produced as *uh* and not as *um*. This finding is somewhat surprising in light of Clark & Fox Tree's (2002) claims that *um* signals that the speaker is having difficulty planning what they want to say while *uh* tends to signal that the speaker is having difficulties in lexical access. Given the complexity of legal discourse and familiarity of both justices and lawyers with legal terminology and jargon, one might expect fewer cases of lexical access difficulties. On the other hand, *uhs* are typically shorter and precede shorter silent pauses than *ums* (e.g. [9]. Hence, greater frequency of *uhs* might follow from the abovementioned strict time restrictions and great time pressure on the speakers.

We next examine the distribution of our four CF types for eight justices. Fig. 1 shows the frequencies for each type as overall CF rate for each justice. We exclude Justice Thomas, who produced only seven CFs in the entire term, compared with the other eight justices, who produced between 308 (STEV) and 1802 (SCAL). It is well known that Justice Thomas speaks very little during oral arguments and the number of fillers (7) strongly correlates with his amount of speech (303 total words).

Table 1
Counts and rates of conversational fillers (CFs).

Conv. filler	Count	CF-frequency	Word-frequency	CF-rate justices	CF-rate lawyers
uh	11,935	67.0	1.53	1.59	1.49
um	2529	14.2	0.32	0.16	0.42
ah	1744	9.7	0.22	0.18	0.25
eh	1598	9.0	0.20	0.23	0.19
misc.	91	0.01	0.0	NA	NA
Total	17,897	100	2.29	2.17	2.37

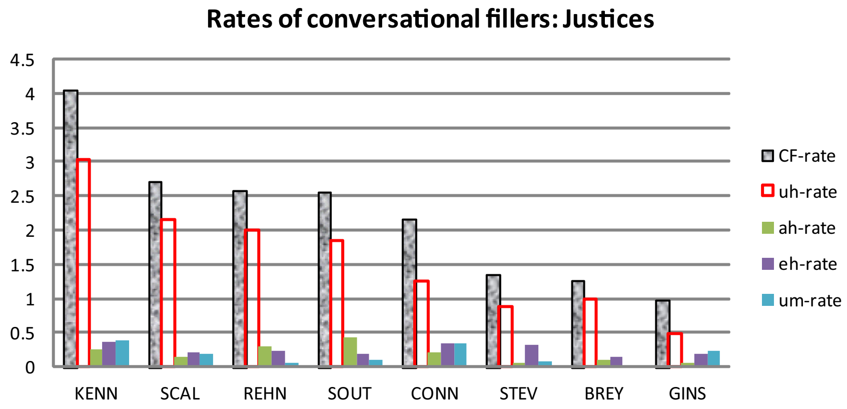

Fig. 1. Distribution of frequencies of four major conversational filler types together with total CF-rate for individual justices.

If the proposal of Clark and Fox Tree [23] about the difference between *um* and *uh* applies to our corpus as well, one might expect lawyers to produce *ums* more frequently than *uhs* than justices, since lawyers are being questioned during the course of their arguments and must provide credible answers quickly. The fifth and sixth columns of Table 1 present CF type rates separately for justices and lawyers. We see that, indeed, the greatest difference between the two groups is in *um*-rate: lawyers produce a higher ratio of *ums* to *uhs* than justices do. A chi-square test with five categories of words (*uh*, *um*, *ah*, *eh*, *other*) shows a significant difference between lawyers and justices; $X^2[4] = 452.6$, $p < 0.001$. A similar result was obtained from a chi-square test with three categories ({*uh, ah, eh*}, *um,* and *other*). Since lawyers use nasalized CFs more often than justices, this corpus provides some support for cognitively different functions of the two main types of CFs.

With respect to the accuracy of the corpus transcription of CFs, Fig. 2 illustrates the vowel quality of all CFs longer than 40 ms, as produced by the eight justices (excluding Thomas) in the F1–F2 space, using the Bark scale [75] for normalizing speaker differences. The first and second formant values (F1 and F2) of the acoustic signal were extracted from the midpoint of the interval aligned to each CF longer than 40 ms to minimize spurious data points. Praat scripts [11] were used for this extraction. Although the plots show considerable amounts of overlap, we see that the Scotus Corpus annotators' transcriptions of the three non-nasalized filler types (*ah, uh, eh*) correspond approximately to their quality. Fillers transcribed as *eh* (depicted in blue) are, in general, more front and high than those transcribed as *uh* (in black) for all justices, while those transcribed as *ah* (in green) are somewhat lower, at least for some speakers. These continuous differences in measured formant values correspond to the discrete differences represented by canonical vowels for these different CFs. We also observe that the vowel quality of *um* (red) is in general very similar to the vowel qualities of *uh* and *ah*. Finally, we also see that the Bark-scale normalization is successful in dealing with gender differences since the values for two female justices (CONN, GINS) are comparable to the male values. We next examine measures of entrainment in the Corpus which explore the use of Corpus transcriptions as well as more objective measures of vowel quality.

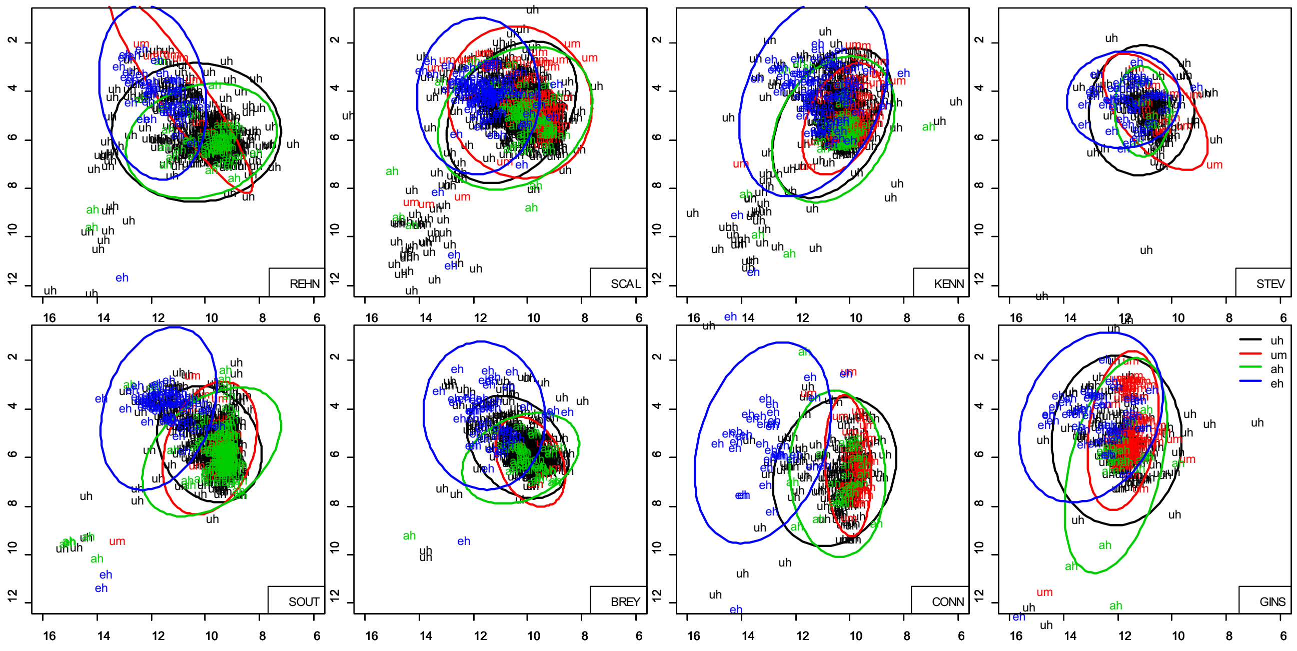

Fig. 2. Vocalic quality of conversational fillers in two-dimensional space of Formant 1 (*y*-axis) and Formant 2 (*x*-axis) Bark values extracted at the midpoint of each filler for eight justices. Ellipses show the 95% confidence intervals.

3.1.2. Entrainment in conversational fillers

To assess the relationship between the quality of the vocalic element of CFs produced by justices and lawyers and to determine if the justices' votes have an effect on this relationship, we employed three of the measures of entrainment described in Section 2.2. We first employ ENTR1 to measure the similarity in the frequencies of usage of CFs in all justice–lawyer pairs. Using the transcripts, we obtained the frequencies of CF types and word counts separately for each oral argument, and calculated entrainment in filler type frequency as in (2).

$$ENTR1(cf) = -\left|\frac{count_{S1}(d)}{ALL_{cf}} - \frac{count_{cf}(d)}{ALL_{S2}}\right| \qquad (2)$$

Here, *cf* represents a particular type of filler (*uh*, *um*, *eh*, etc.); *S*1 and *S*2 represent a justice–lawyer pair of speakers in an oral argument; $count_S(cf)$ is the number of times speaker *S* used *cf*, and ALL_S is the number of all words uttered by *S*. Entrainment is thus the difference between the filler frequency in their speech.

The limitation of this static measure lies in the fact that the transcriptions were performed by multiple persons and additional processing of the transcripts, which included inserting omitted fillers, was performed by another group of coders. Since no information on the inter-labeler agreement is available, it is plausible that discrete labels of CFs (*uh*, *um*, *ah*, *eh*) were not applied consistently.[1]

We employ two other measures of entrainment, ENTR2 and ENTR3, to identify similarities in the continuous information of vowel quality of Cfs used by justices and lawyers, which we calculate from the first (F1) and second formant (F2) values for CF vowels. Given the means of F1 and F2 for each speaker in each oral argument, we examine the Euclidean distance between these means for each justice–lawyer pair. Hence, these distances give us another global measure of entrainment between a lawyer and a justice within a session: If the Euclidean distance is smaller in the justice_x–lawyer_y pair than in the justice_x–lawyer_z pair, we say that lawyer_y is more similar to justice_x than lawyer_z is, in terms of the quality of the vocalic element of a CF.

We use our third measure of entrainment, ENTR3, to assess the similarity of lawyer-justice CFs more locally. The oral arguments frequently include multiple short dialogs between a lawyer and a justice containing multiple turns. ENTR2 would not detect whether a lawyer adjusts their speech to a particular justice in such local dialogs. We explore this type of CF entrainment by calculating the distance between lawyer-justice CFs for only adjacent CFs for the lawyer-justice pair. This measure gives us Euclidean distance values for each pair of adjacent CFs between a lawyer and a justice. We illustrate this measure in Fig. 3 below where each line represents one value of our dependent variable ENTR3, corresponding to the Euclidean distance between the F1 and F2 values of the two CFs at the two ends of the line segment.

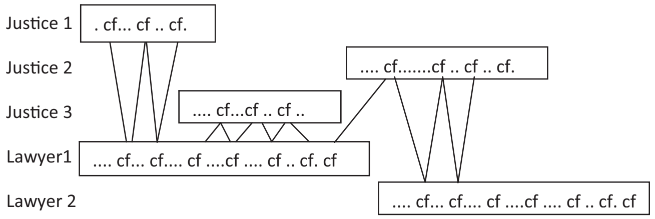
Justice 1
. cf... cf .. cf.
Justice 2
.... cf.......cf .. cf .. cf.
Justice 3
.... cf...cf .. cf ..
Lawyer1
.... cf... cf.... cf ....cf .... cf .. cf. cf
Lawyer 2
.... cf... cf.... cf ....cf .... cf .. cf. cf

Fig. 3. Idealized speech (represented as boxes) of justices and a lawyer with interspersed conversational fillers (cf).

Since the publicly available annotation database of Supreme Court hearings allows for determination of justice voting on a case and association of lawyers to these cases, we split all justice–lawyer pairs into two groups (justice voted in favor of the lawyer vs. against). For each entrainment measure we then test for the effect of the justices' decisions in favor or against the case presented by the lawyers in the session.

We perform a series of t-tests to compare the two justice VOTE groups (in-favor vs. against the lawyer's case) using each of our measures of entrainment of CFs in turn as dependent variables. Recall that our first measure, ENTR1, measures the difference in CF rates per CF type and per session in lawyer-justice pairs. For this simple measure, we find that whether a justice voted for or against a lawyer's case has no significant effect on similarity in CF rates between lawyer and justices (t = 0.24, df = 1401, p = 0.81). When we examine whether individual justices conformed to this general result, t-tests again show no significant effect of voting behavior on ENTR1. In addition, the polarity of the t statistic was evenly split (4–5), so we do not even find that the direction in which an individual justice tended in their vote was influenced by CF entrainment measured under ENTR1.

Our second metric, ENTR2, attempts to capture CF entrainment in terms of vowel quality – similarity in F1–F2. When we perform similar t-tests with this measure as the dependent variable, we find that mean distances for lawyer-justice pairs in which the justice voted for a lawyer's position were indeed smaller than those when the justice voted against that position. However, this effect was small and not statistically significant (t = 1.25, df = 1081.5, p = 0.21). Again testing for individual differences among the justices, we found no difference at alpha = 0.05 and only one result approaching significance at alpha = 0.1 (RHEN). We found that six of the eight justices were more similar in CF quality to the lawyers they voted ***for*** (including the one justice whose similarity approached significance at p < 0.1), while two justices produced CFs more similar to the lawyers they voted ***against***. Hence, while we find tendencies for entrainment between lawyers and justices using this metric, albeit in different directions, the differences are not statistically significant.

In our final set of t-tests, we compared justice voting behavior against our third measure of CF entrainment, ENTR3. Recall that this metric tests CF production similarity between ***adjacent*** CFs in lawyer-justice turns. This local entrainment metric ***does*** show a significant effect: mean distance between adjacent CFs was smaller for lawyer-justice pairs in which the justice gave a favorable vote than in those pairs in which the justice gave vote against the lawyer (t = 2.26, df = 982.1, p = 0.024). We found similar results when we tested separately for lawyers representing the petitioner's side (t = 1.98, df = 432.2, p = 0.049). In separate tests for individual justices, one test showed significance (t = 2.13, p = 0.035, GINS) and one tendency (t = 1.75, p = 0.084, SOUT) in the same direction as the overall t-tests. In these tests, seven t-values were positive (including the two already mentioned above) and only one was negative (although not significant). These findings suggest that, if a lawyer produces CFs similar to the justice s/he ***currently*** is speaking with during the oral argument, this justice will indeed be likely to vote in favor of that lawyer's side of the case.

Our findings with respect to the influence of CF entrainment on justices' voting behavior, while intriguing, must nonetheless be viewed with caution. First, this is a first preliminary investigation applying most rudimentary measures and analyses. Second, CF similarity might be related to phonetic context, for which we did not control. For example, if one speaker tends to produce CFs following *the*, their quality will be influenced by this preceding vowel. So, the entrainment between speakers that we see might be due to features other than the phonetic ones we are considering. Third,

[1] However, this problem is alleviated by the nature of our analysis, since the argument of each docket was presumably labeled consistently and our analysis compares language within and not across dockets.

the directionality in adjacent CFs might have an effect. Our third measure was bi-directional; hence, a CF from a lawyer could have been influenced by a preceding CF from a justice as well as the justice's following one. Fourth, bi-directionalilty of this measure cannot determine if lawyers entrain to justices or vice versa; but see the following section for some indication for the former in intensity. Finally, justices base their decisions on many other aspects besides the oral arguments and lawyers are aware of their leanings before the oral argument [27]. So, additional investigations must be made to assess the actual contribution of this form of lawyer-justice entrainment.

3.2. Local Intensity Entrainment and Power Relations

As discussed in Section 1, entrainment between conversational partners is commonly associated with dominance. When power imbalance exists between interlocutors, the less dominant speaker will converge more [33]. Furthermore, a speaker should entrain more to an interlocutor whom she likes than to one whom she dislikes. We therefore predict that in our domain, lawyers, as lower-status interlocutors, should entrain more to justices than justices to lawyers, and justices should entrain more to lawyers for whom they ultimately vote. Similar hypotheses were confirmed on this data for linguistic features [27].

Since females are known to have greater perceptual sensitivity to vocal characteristics, they are likely to entrain more to such characteristics. However, experimental results testing this hypothesis have been mixed [10,59,51,49]. Here, we also look at differences in entrainment between male and female lawyers.

We looked at differences in entrainment on vocal intensity using our fourth measure of entrainment, ENTR4, which describes the extent to which $S1$ matches $S2$ when beginning a new turn. Recall that IPUs are defined as pause-free segments of speech from a single speaker, surrounded by silence longer than 50 ms. For each turn in the corpus, we compared the first IPU of that turn with the last IPU of the preceding turn, taking the negated absolute value of the difference between the two segments' intensity values.

Our analysis showed that lawyers do in fact entrain to justices more than justices entrain to lawyers. Turns belonging to lawyers were significantly more similar in intensity to their preceding turns than turns belonging to justices ($t = 7.02$, df = 17622, $p < 0.001$, mean lawyer similarity = −3.59, mean justice similarity = −3.95). This finding supports the hypothesis that a less dominant interlocutor is likely to entrain more than a dominant one. However, we did not find a significant difference in entrainment between male and female lawyers ($t = 1.29$, df = 2205.1, $p = 0.20$, mean male similarity = −3.61, mean female similarity = −3.50).

Our results comparing entrainment between justices and the lawyers with whom they do or do not side in their decision were mixed. We found that differences between justices and petitioners were significantly smaller when the justice sided with the petitioner ($t = -2.14$, df = 294.86, $p = 0.03$, mean petitioner similarity = −3.71, mean respondent similarity = −4.18). However, differences between justices and respondents were also significantly smaller when the petitioner won the case ($t = 2.53$, df = 217.9, $p = 0.01$, mean petitioner similarity = −3.68, mean respondent similarity = −4.26). In other words, justices entrained more in general whenever the petitioner ultimately won the case, independent of whether the justice voted for the petitioner or the respondent.

In general, our results support the theories of entrainment and dominance that predict that the less dominant speaker will entrain more. However, we do not find that justices converge more to the lawyer with whom they eventually side; nor do we find that females converge significantly more.

3.3. Turn-taking

Speakers' production of CFs is closely linked to turn-taking behavior, since a primary function of CFs is to signal the interlocutor that the speaker wishes to hold the floor or to take the floor or to relinquish it or to acknowledge the need for information [71,8]. We noted above that our corpus has been annotated for overlaps between speakers. With this information, we calculate a set of TURN-TAKING PATTERNS for justices and lawyers, identified by examining sequences of justice–lawyer speech and the presence or absence of overlap between the turns, as follows:

- J-L: Speech segment from a justice followed by a segment from a lawyer; no overlap.
- L-J: Speech segment from a lawyer followed by a segment from a justice; no overlap.
- J-JL: Speech segment from a justice, followed by a segment with overlap from the same justice and a lawyer.
- L-JL: Speech segment from a lawyer, followed by a segment with overlap from the same lawyer and a justice.
- JL-J: Speech segment with overlap from a justice and a lawyer, followed by a segment from the same justice.
- JL-L: Speech segment with overlap from a justice and a lawyer, followed by a segment from the same lawyer.

These patterns were examined with respect to their frequency of occurrence when the justice in the exchange voted against or for the lawyer in the exchange (VOTE factor).

The figure below presents contingency tables showing the frequencies of each pattern with respect to whether justices voted for or against the lawyer participating in the exchange. Fig. 4 thus shows a contingency table for each individual justice (again, Thomas was excluded due to data scarcity). In the figure, the width of the columns is proportional to the frequency of the turn-taking patterns. Each column is divided into two parts, according to the proportion of votes in favor or against.

The figure shows a few interesting tendencies in terms of frequencies of turn-taking patterns. For example, for all judges, "J–JL" and "JL–L" are the least frequent transition patterns, suggesting that lawyers are unlikely to initiate or win an overlap. Further, all judges but REHN show a higher frequency of "L–JL" than "L–J", which indicates that lawyers' turns are more likely to end in a judge-initiated overlap.

Additionally, in these plots, if Vote were not a factor, they would show an approximately horizontal line – that is, a similar proportion of each pattern relative to in-favor vs. against votes. For justices Breyer and Rehnquist, a Chi-Squared test reveals a significant departure from that case ($p < 0.01$); for O'Connor and Stevens, the same result approaches significance ($p < 0.1$). This constitutes evidence of an effect of justice decision on turn-taking behavior, at least for some of the justices.

By observing each justice's contingency table, we may examine how their turn-taking behavior differs relative to their vote. In Justice Breyer's case, favored lawyers appear more likely to both initiate a speech overlap and continue speaking after one (third and fourth columns in the corresponding figure). Justice Rehnquist seems more likely to initiate speech overlaps and continue speaking after one (fifth and sixth columns).

However, it should be noted that from the turn-taking annotations used in this study, it is difficult to predict the type of exchange that actually took place. For example, an "L–JL" transition may correspond to a short overlap (J starts speaking during the very end of L's contribution, without interrupting) or to an interruption (J starts speaking before L can complete their utterance, thus interrupting). Therefore, to better understand these results we need to analyze overlapping transitions in more detail. For that

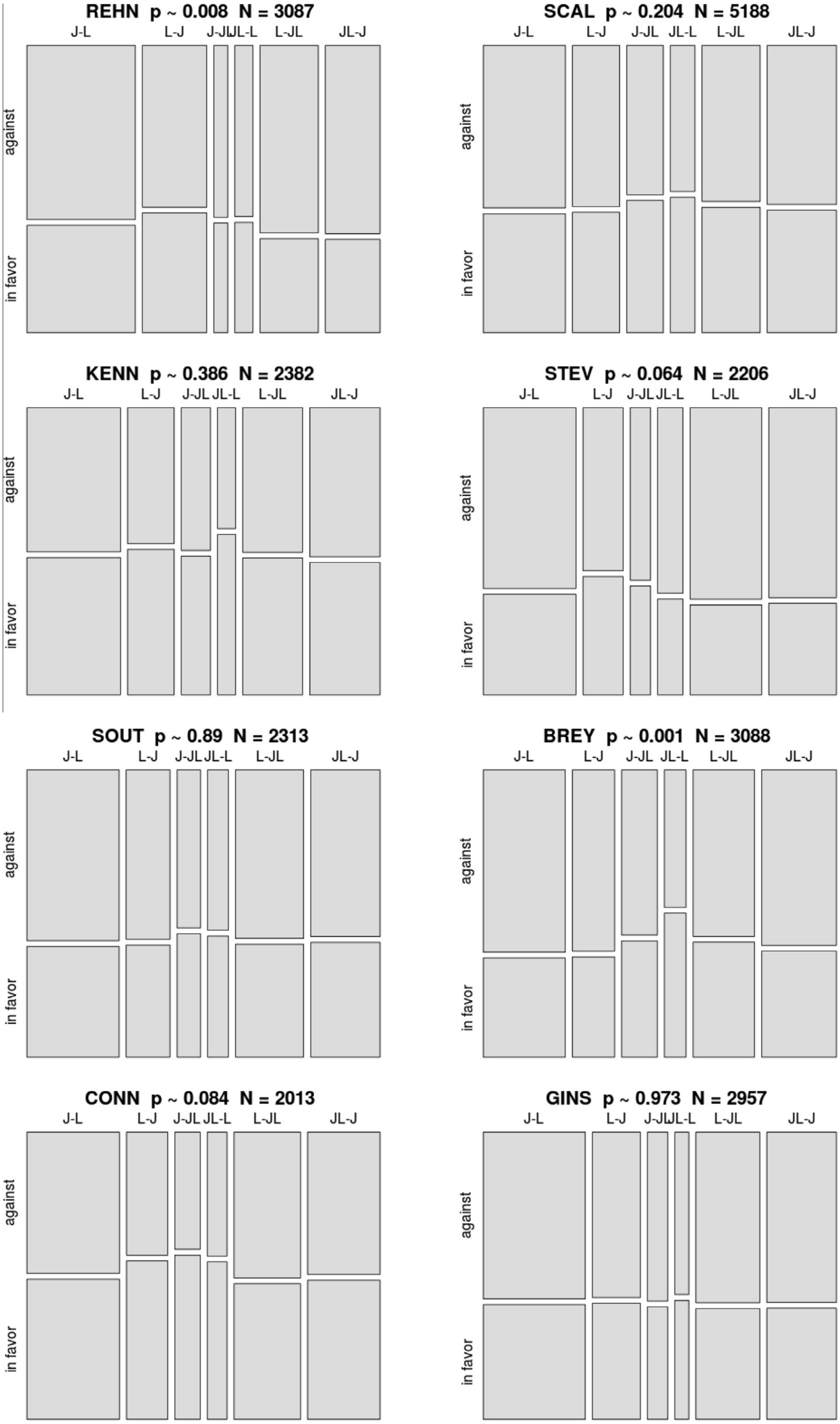

Fig. 4. Visual representation of a contingency table for VOTE vs. TURN-TAKING PATTERN, considering each Justice individually.

purpose, we proceeded to manually annotate a portion of the corpus for type of overlapping transitions – more specifically, we focused on transitions in which speaker *S*1 was holding the turn, and speaker *S*2 started speaking, overlapping *S*1's speech.

We automatically sampled 960 overlap instances from the corpus, using the dialogue transcripts as a reference. These instances were balanced for justice identity (eight in total, after excluding Thomas), justice vote (in favor of or against the lawyer's case) and overlap pattern. We considered four possible overlap patterns, according to the dialogue transcripts: J–JL–J (meaning, a speech segment from justice J, followed by an overlap between J and lawyer L, followed by a segment from J), J–JL–L, L–JL–J and L–JL–L. According to the dialogue transcripts, these four patterns cover all possible speech overlaps between exactly one justice and one lawyer.

This labeling scheme is an extension of one we have used in previous studies (e.g. [37]. It identifies the following kinds of overlap:

- **BC** Backchannel: *S*2's utterance is in response to *S*1's utterance and indicates only "I'm still here/I hear you and please continue".
- **O** Overlap: *S*1's utterance is almost complete at the time *S*2 starts; *S*2 successfully takes the turn; the overlapping speech starts when *S*1 is almost done speaking (i.e., *S*1 is completing the last few syllables of his/her intonational phrase).
- **LO** Long overlap: Same as O, but the overlap spans a longer speech segment.
- **EO** Embedded overlap: A short, complete turn like 'no' or 'that's correct' while the current speaker holds the turn.
- **I** Interruption: *S*1's utterance is incomplete at the time *S*2 starts; *S*2 successfully takes the turn.
- **BI** Butting-in: *S*1's utterance is incomplete at the time *S*2 starts; *S*2 does not manage to take the turn (*S*1 continues speaking).

Two trained annotators first labeled 100 samples separately, and achieved a Kappa measure of 0.714, which is considered a 'substantial' degree of agreement [24]. Given this degree of agreement on the first sample, each annotator then labeled half of the remaining samples. As a result, 29 samples were labeled BC, 142 O, 78 LO, 96 EO, 257 I, and 170 BI. Also, 188 samples were discarded, because of errors in the transcripts or because they did not actually match our definitions of overlap types (for example, several cases corresponded to simultaneous starts by two speakers after a long silence).

For reference, we compare the distribution of overlap types in these samples from the Scotus Corpus with that of the collaborative task-oriented dialogs in the Columbia Games Corpus (CGC) [37]. We observe a lower rate of backchannels (Scotus: 4%; CGC: 13%), a lower rate of overlaps (Scotus: 41% after collapsing O, LO and EO; CGC: 71%), and a higher rate of interruptions (Scotus: 33%; CGC: 10%) and butting-ins (Scotus: 22%; CGC: 7%). Further, in practically all instances of the overlap (O) category in the CGC, the overlap duration is shorter than one second, while in the Scotus Corpus, the mean overlap duration is 1.15 s (SD = 0.77). These pronounced differences appear to reflect the different characteristics of the two corpora: the CGC consists of collaborative, low-stakes conversations of people playing simple computer games in a relaxed setting; the Scotus Corpus consists of high-stakes legal disputes taking place in a time-constrained scenario. For each justice, and for all justices together, we computed two contingency tables – one for the case in which the justice was holding the turn and the lawyer began to speak, and the other for the inverse case. Initially, each table contained two rows (voted against/for lawyer) and six columns (one for each turn-taking label). Given that several cells contained small values, we collapsed two pairs of labels that corresponded to similar turn-taking categories: We collapsed embedded overlaps (EO) and backchannels (BC); in fact, most instances of embedded overlaps were merely short acknowledgments, which are pragmatically close to backchannels. We collapsed overlaps (O) with long overlaps (LO); the only difference between these two categories is the duration of the overlapping segment. The resulting contingency tables thus contain two rows and four columns (O, EO, I, BI). We performed Fisher's Exact Tests on these tables, searching for evidence of deviation from a random distribution – that is, that there is an effect of justice vote on turn-taking behavior.

For the case in which the lawyer held the turn and the justice produced an overlap, we found a number of results approaching significance (Fisher's Exact Test, *p*-value < 0.1), as shown in the following tables.

All justices together (*p*-value = 0.082):

	O	**EO**	**I**	**BI**
Against	43	32	95	46
In favor	42	39	114	28

Justice Breyer (*p*-value = 0.077):

	O	**EO**	**I**	**BI**
Against	3	6	13	4
In favor	9	6	16	0

Justice Souter (*p*-value = 0.074):

	O	**EO**	**I**	**BI**
Against	9	4	8	7
In favor	2	5	15	5

According to our labeling scheme, an interruption attempt may lead either to a successful interruption (I) or to an unsuccessful one (BI). The results depicted above suggest that, when a lawyer is speaking and a justice attempts to interrupt, it is more likely for the interruption to succeed *when the justice is favorable to the lawyer's case*. In other words, lawyers seem more likely to yield to an interruption attempt when the interrupting justice is favorable to their case than otherwise. For the inverse case, in which the justice held the turn and the lawyer produced an overlap, we found no statistically significant differences. These observations, however, need to be further examined in view of the finding that the more a justice talks to or questions the lawyer, the less likely s/he is to vote for the lawyer [68].

3.4. Lexical cues to judicial alliance

In this section, we examine whether there are lexical cues indicating degree of agreement of the Justices themselves on a decision. We compare close verdicts (defined as being split with five Justices voting one way, and four voting the other) and unanimous verdicts, using Linguistic Inquiry and Word Count (LIWC; [57]. LIWC is a tool used to analyze words in text according to 74 different categories. For each category, LIWC returns the occurrence of that category in the sample as a percentage of total words. The categories range from purely grammatical (i.e. measuring use of articles, personal pronouns, negations) to more abstract categories such as *positive emotion words*, *insight words*, and *social words*. These categories were initially formed from judgments by independent judges. LIWC has been used to find indicators of deception [56], marital happiness [70], and dominance relations [66]. We performed *t*-tests to determine which categories most significantly differed between the two categories, and then used the results as the basis for feature selection in a classification experiments designed to predict whether a case will be close.

Politically, there was a 5–4 Republican–Democrat split among the Justices in the 2001 term, which led to many close cases. Of the 78 cases analyzed, 24 were decided with a 5–4 split and 27 were unanimous. Overall the justices uttered 350,855 words, but these were by no means divided equally between all of them. The average was just under 40,000 words per justice; however, Justice Thomas spoke only 303 words the entire year, whereas Justice Breyer spoke 103,932 words over the course of the year.

We originally sought a binary labeling of "close cases" and "unanimous cases", but the way justices decide cases makes that difficult. A justice who agrees with the majority opinion delivers the Court's opinion, and a justice participating in the minority opinion files a dissent. The rest of the justices who agree with the majority can either join the Court's opinion, or file a concurrence. A concurrence is a document that supports the general opinion of the majority, but disagrees with either the reason for the decision, or a minor point of the decision. Justices who agree with the minority opinion can similarly file dissenting concurrences. We collapsed concurrences and dissenting concurrences into their respective opinion so as to obtain a binary labeling. In one case, a justice filed both a concurring and a dissenting opinion and this was discarded due to ambiguity.

For each category we took the LIWC output (the prevalence of a category as a percentage of words used from that category divided by total words used) and normalized using *z*-score normalization. We then performed t-tests to find factors that appeared to contribute to a close decision. Of the significant differences we found, the most significant ($p \leqslant .005$) is that word count ($t = 3.08$, df = 48.8, $p = 0.003$) and the use of present tense verbs ($t = 2.94$, df = 41.8, $p = 0.005$) are greater in close cases. Sessions that were close cases had on average 860 more words per session than unanimous cases, a difference of 8.3%.

We used the LIWC features in classification experiments using decision trees and logistic regression to predict the agreement of the justices – that is, whether a case was split or unanimous. We used the significance of each LIWC category in the *t*-test for feature selection, and classified the cases using ten-fold cross-validation using only the categories whose *p*-value was below a certain threshold. The baseline accuracy assumes always choosing the majority class (unanimous). The results are summarized in Table 2.

Present tense verbs and word count (the features with $p \leqslant .005$) had the best classification accuracy for logistic regression at 62.4%, whereas word count as the sole classification feature ($p \leqslant .003$) had the highest classification accuracy for decision trees at 74.1%. The more words used by justices in the oral session, the more likely it was that the case would be a close decision. This intuitively makes sense, as justices would be likely to ask more questions in a case which is harder to decide than in a case that is easy to decide. The increased use in present tense verbs in split cases was also a very strong indicator of how close a case was. When discussion is focused on the present activities as opposed to precedent and past cases, this can indicate that the potential outcome of a case is relatively subjective, and therefore more likely to be split along ideological lines or affected by power dynamics.

The other LIWC dimensions significantly associated with a close decision ($p < .05$) are increased use of the first person singular pronouns ($t = 2.14$, df = 40, $p = 0.037$), and increased use of verbs ($t = 2.09$, df = 43.7, $p = 0.042$). An increase in the frequency of first person singular pronouns (e.g. I, I'm, me) would indicate that the discussion involves more personal opinions, and is therefore likely to be more contentious. An increased use of verbs in general was a significant feature, likely due to the fact that increased present tense verbs was an even better predictor of close cases. However, adding the use of first person singular pronouns and verbs as features did not improve upon the classification accuracy of using the most significant features, although they were still an improvement over the baseline. Including all the features with *p*-values less than 0.1 saw a decrease in accuracy across both classifiers, along with using all the features regardless of *p*-value. Word count and present tense verbs remained the most significant features as well as the best predictors of the agreement of the justices.

Table 2
Accuracy of agreement classification for different features.

Significance threshold	Logistic regression	Decision trees
$p \leqslant .003$	60.6	74.0
$p \leqslant .005$	62.4	63.7
$p \leqslant .05$	60.0	61.5
$p \leqslant .1$	56.8	57.4
All	37.0	49.4
Baseline	52.9	52.9

Features with $p \leqslant .003$: Word Count; Features with $p \leqslant .005$: Word Count, Present Tense Verbs; Features with $p \leqslant .05$: Word Count, Present Tense Verbs, First Person Singular Pronouns, Verbs; Features with $p \leqslant .1$: Word Count, Present Tense Verbs, First Person Singular Pronouns, Verbs, Six Letter Words, Dictionary Words, Function Words, Past Tense, Numbers, Positive Emotion, Discrepancy Words, Tentative Words.

4. Discussion

The goal of this paper has been to investigate the role of entrainment in the production of conversational fillers and acoustic intensity; patterns of turn-taking; and Linguistic Style markers as communicative social signals related to power relations, conflict, and voting behavior. We examine these issues in a corpus of speech produced by justices and lawyers during oral arguments of the U.S. Supreme Court in the 2001 term. We examined three possible measures of entrainment between justices and lawyers to see whether they were related to justices' favorable or unfavorable votes for the lawyers. Two tests – a naïve measure of similarity in conversational filler rates and global similarity in filler's vowel quality for the entire session – showed no relationship. The third, which measured local entrainment in conversational fillers in lawyer-justice pairs, did in fact identify a significant positive relationship between entrainment and justice votes. With respect to local entrainment in intensity, we found that lawyers did entrain more to justices than justices to lawyers; female lawyers did not entrain any more than did male lawyers. When we examine the relationship between entrainment in intensity and judicial voting, we found that when justices voted for the petitioners, there was significant evidence of a greater degree of entrainment by both petitioners and respondents to justices. With respect to turn-taking behavior, we have found that certain patterns of overlaps in turn exchanges between justices and lawyers were correlated with justices' voting behavior for four of the justices in our corpus. A more detailed examination of a sample of our data shows that, when a lawyer is speaking and a justice attempts to interrupt, the interruption is more likely to succeed when the justice is favorable to (i.e. subsequently votes in favor of) the lawyer's side of the case. Finally, we have examined whether there are lexical cues to divisiveness within the Court itself on a particular case: that is, whether there are lexical indicators that distinguish cases with close verdicts from cases with unanimous verdicts. Using the LIWC tool, we have identified a number of characteristics that distinguish the two situations: close decisions are characterized by greater number of words spoken by the justices, greater use of the first person singular, greater number of verbs used and, in particular, an increase in the present tense. While greater number of words spoken seems quite plausible in close cases, and greater use of "I" may signal an increase in the expression of personal opinions, greater use of the present tense is not so readily explained: perhaps less availability of or reliance on past precedent is in fact what characterizes these cases.

Several aspects of these results are relevant for the emerging field of Cognitive Info-Communications (CogInfoCom). One of the primary aims of this research program is to facilitate building interfaces that bridge the cognitive capabilities of humans with those of artificial systems, and possibly even extend the human ones. We argue that entrainment between humans, which is afforded by the spoken modality of interaction through the acoustic channel, represents a fundamental cognitive ability of humans. We believe that understanding how this ability is deployed for dynamic negotiations of social relationships, such as evolving dominance asymmetries or building alliances in multi-party interactions, represents one of the key building blocks for producing future effective interfaces for human–machine interactions. Moreover, the usefulness of such interfaces will be greatly extended if we understand how parameters such as *stakes in the outcome* or level of emergency affect the deployment of entrainment as a communicative social signal.

We see our contribution towards these questions to be twofold. First, we show that in high-stakes situations, degree of entrainment correlates with (and possibly has an effect on) the outcome. Crucially, this correlation obtains mostly for local measures of entrainment. For potential applications, this highlights the importance of dynamic, online adjustments made on a turn-by-turn basis rather than modifications of global settings (see also [48] for similar findings with a different corpus). This observation builds groundwork for future exploration of dynamic fluctuation of the relationship between entrainment on the one hand and dominance and alliance on the other hand. Our results also support the idea that communicative social signals negatively correlate with the semantic load of utterances since vocal intensity and quality of conversational fillers have minimal linguistic functions, and temporal aspects of turn-taking behavior, i.e. when a speaker initiates speech, are also meaningful pragmatically rather than semantically. Hence, utilizing entrainment in the architecture of CogInfoCom interfaces opens the possibility for engineering applications that effectively incorporate social skills into human–machine interactions. Moreover, CogInfoCom interfaces in these applications might facilitate communication between humans and systems by using observed predictive patterns in the cognitive system that links speaking behavior with social structure. For example, using the tendencies reported in our turn-taking experiment, if a higher-status agent (human) tolerates interruptions from a lower-status agent (system), this might predict a favorable outcome (by the human). In the absence of such toleration, the system might decide to entrain even more to the human on other dimensions (e.g. intensity or vowel quality). Furthermore, recognizing and producing some entrainment features, such as conversational fillers, or intensity, are relatively easy tasks from the engineering point of view.

The second aspect of our results that is linked to the CogInfoCom research agenda relates to the question of how entrainment could be sensed by the system; in other words, can the production–perception loop of the system access and evaluate features of entrainment directly from the acoustic signal, or does the system need to infer some representational categories, shared with the human, for the production and evaluation of entrainment. This relates to the yet unresolved debate whether entrainment is a fast, low-level and largely mechanistic feature of human–human interactions or if it requires higher-level cognitive processes (c.f. [60]. Our results suggest that entrainment is a multi-dimensional feature and that some aspects can be extracted directly from the signal, such as intensity of speech or temporal features of turn-initiation, while some might require detecting some relatively low-level categories, such as binary decision if a conversational filler is present or not. However, some aspects might require more complex categories such as parts of speech or even a very complex inference model. Consider for example the results obtained from comparing close vs. unanimous decisions: Justices do not talk to each other during oral arguments, yet they use their interactions with lawyers for negotiating alliances among themselves [31]. Given certain, and likely very different, models of social structure incorporating the notions of dominance and status between humans and system, a CogInfoCom interface between these models using entrainment requires a broad spectrum of perception–production mechanisms with varying degrees of cognitive complexity.

In sum, this paper suggests that entrainment in spoken interactions, and the patterns for its utilization for negotiating social relations, represent a fruitful field for building formal models and engineering applications that facilitate more effective interactions between human and artificial cognitive systems.

Acknowledgments

This work was supported in part by the National Science Foundation under Grant No. IIS-0803148, ANPCYT PICT-2009-0026, CONICET and by the project implementation: Technology research for the management of business processes in heterogeneous distributed systems in real time with the support of multimodal communication, ITMS 26240220064 supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] M. Babel, Evidence for phonetic and social selectivity in spontaneous phonetic imitation, J. Phonet. 40 (2012) 177–189.

[2] M. Babel, D. Bulatov, The role of fundamental frequency in phonetic accommodation, Lang. Speech 55 (2) (2011) 231–248.

[3] P. Barányi, G. Persa, A. Csápó, Definition of cognitive infocommunications and an architectural implementation of gognitive gnfocommunications systems, World Acad. Sci. Eng. Technol. 58 (2011) 501–505.

[4] L.M. Beebe, Social and situational factors affecting the communicative strategy of dialect code-switching, Int. J. Sociol. Lang. 32 (1981) 139–149.

[5] L. Bell, J. Gustafson, M. Heldner, Prosodic adaptation in human–computer interaction, in: Proceedings of 15th International Congress of Phonetic Sciences, 2003.

[6] Š. Beňuš, Social aspects of entrainment in spoken interaction. Cogn. Comput. (in press) http://dx.doi.org/10.1007/s12559-014-9261-4.

[7] Š. Beňuš, Variability and stability in collaborative dialogues: turn-taking and filled pauses, in: Proceedings of the 10th INTERSPEECH, 2009, pp. 709–799.

[8] Š. Beňuš, A. Gravano, J. Hirschberg, Pragmatic aspects of temporal entrainment in turn-taking, J. Pragmat. 43 (12) (2011) 3001–3027.

[9] S. Benus, F. Enos, J. Hirschberg, E. Shriberg, Pauses and deceptive speech, in: Proceedings of 3rd International Conference on Speech Prosody, 2006.

[10] F. Bilous, R.M. Krauss, Dominance and accommodation in the conversational behavior of same- and mixed-gender dyads, Lang. Commun. 8 (1988) 183–194.

[11] P. Boersma, D. Weenink, Doing phonetics by computer, 2013. (computer program, www.praat.org).

[12] R.Y. Bourhis, H. Giles, W.E. Lambert, Social consequences of accommodating one's style of speech: a cross-national investigation, Int. J. Sociol. Lang. 6 (1975) 55–72.

[13] R. Bourhis, H. Giles, The language of intergroup distinctiveness, in: H. Giles (Ed.), Language, Ethnicity and Intergroup Relations, Academic Press, London, 1977, pp. 119–135.

[14] S.E. Brennan, Lexical entrainment in spontaneous dialog, in: Proceedings of the International Symposium on Spoken Dialog (ISSD), 1996.

[15] S.E. Brennan, H.H. Clark, Conceptual pacts and lexical choice in conversation, J. Exp. Psychol.: Learn. Memory Cogn. 22 (6) (1996) 1482–1493.

[16] S.E. Brennan, J.E. Hanna, Partner-specific adaptation in dialogue, Top. Cogn. Sci. 1 (2009) 274–291.

[17] S.E. Brennan, M. Williams, The feeling of another's knowing: prosody and conversational fillers as cues to listeners about the metacognitive states of speakers, J. Memory Lang. 34 (1995) 383–398.

[18] E.H. Buder, A.S. Warlaumont, D.K. Oller, L.B. Chorna, Dynamic Indicators of Mother–Infant Prosodic and Illocutionary Coordination, in: Proceedings of 5th Speech Prosody, 2010.

[19] D.B. Buller, R.K. Aune, The effects of vocalics and nonverbal sensitivity on compliance: a speech accommodation theory explanation, Hum. Commun. Res. 14 (3) (1988) 301–332.

[20] J. Burgoon, N. Dunbar, An interactionist perspective on dominance submission: interpersonal dominance as a dynamic, situationally contingent social skill, Commun. Monogr. 67 (2000) 96–121.

[21] J.N. Cappella, S. Planalp, Talk and silence sequences in informal conversationa III: Interspeaker influence, Hum. Commun. Res. 7 (2) (1981) 117–132.

[22] T. Chartrand, J. Bargh, The chameleon effect: the perception–behavior link and social interaction, J. Pers. Soc. Psychol. 76 (1999) 893–910.

[23] H.H. Clark, J.E. Fox Tree, Using uh and um in spontaneous speaking, Cognition 84 (2002) 73–111.

[24] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Measure. 20 (1) (1960) 37–46.

[25] R. Coulston, S. Oviatt, C. Darves, Amplitude convergence in children's conversational speech with animated personas, in: Proceedings of ICSLP, 2002.

[26] A. Csapó, P. Baranyi, A conceptual framework for the design of audio based cognitive infocommunication channels, in: J. Fodor et al. (Eds.), Recent Advances in Intelligent Engineering Systems, Springer-Verlag, Berlin Heidelberg, 2012, pp. 261–281.

[27] D. Danescu-Niculescu-Mizil, L. Lee, B. Pang, J. Kleinberg, Echoes of power: Language effects and power differences in social interaction, in: Proceedings of the 21st International Conference on World Wide Web, 2012, pp. 699–708.

[28] Veronique Delvaux, Alain Soquett, The influence of ambient speech on adult speech productions through unintentional imitation, Phonetica 64 (2007) 145–173.

[29] N.E. Dunbar, J.K. Burgoon, Perceptions of power and interactional dominance in interpersonal relationships, J. Soc. Personal Relat. 22 (2005) 231–257.

[30] N.E. Dunbar, A.M. Bippus, S.L. Young, Interpersonal dominance in relational conflict: a view from Dyadic Power Theory, Interpersona 2 (1) (2008) 1–33.

[31] L. Epstein, W. Landes, R.A. Posner, Inferring the Winning Party in the Supreme Court from the Pattern of Questioning at Oral Argument, University of Chicago Law & Economics, Olin Working Paper No. 466, 2009.

[32] S. Feldstein, C.L. Crown, Oriental and Canadian conversational interactions: chronographic structure and interpersonal perception, J. Asian Pacific Commun. 1 (1990) 247–266.

[33] H. Giles, A. Mulac, J.J. Bradac, P. Johnson, Speech Accommodation Theory: The Next Decade and Beyond, Communication Yearbook, Sage, Newbury Park, 1987, pp. 13–48. 10.

[34] H. Giles, N. Coupland, J. Coupland, Entrainment theory: communication, context and consequence, in: H. Giles, J. Coupland, N. Coupland (Eds.), Contexts of Entrainment: Developments in Applied Sociolinguistics, Cambridge University Press, Cambridge, 1991, pp. 1–68.

[35] D. Goleman, Social Intelligence: the new science of human relationships, Bantam (2006).

[36] D. Gorter, Aspects of language choice in the Frisian-Dutch bilingual context: neutrality and asymmetry, J. Multilingual Multicul. Develop. 8 (1987) 121–132.

[37] A. Gravano, J. Hirschberg, Turn-taking cues in task-oriented dialogue, Comput. Speech Lang. 25 (3) (2011) 601–634.

[38] S.W. Gregory, S. Webster, A nonverbal signal in voices of interview partners effectively predicts communication entrainment and social status perceptions, J. Pers. Soc. Psychol. 70 (1996) 1231–1240.

[39] B. Guitar, L. Marchinkoski, Influence of mothers' slower speech on their children's speech rate, J. Speech Lang. Hear. Res. 44 (2001) 853–861.

[40] J. Hay, S. Jannedy, N. Mendoza-Denton, Oprah and/ay/: lexical frequency, referee design and style, in: Proceedings of International Congress of Phonetic Sciences, 1999.

[41] J. Hirschberg, Speaking more like you: Entrainment in conversational speech, in: Proceedings of Interspeech, 2011, pp. 27–31.

[42] http://scdb.wustl.edu/.

[43] http://www.oyez.org/.

[44] R. Jacobson, Closing statements: linguistics and poetics, in: T.A. Sebeok (Ed.), Style in Language, MIT Press, Cambridge, Massachusetts, 1960, pp. 350–377.

[45] J. Jaffe, S. Feldstein, Rhythms of Dialogue, Academic Press, New York, 1970.

[46] M. Kontra, M. Gosy, Approximation of the standard: a form of variability in bilingual speech, in: A.R. Thomas (Ed.), Methods in Dialectology, Multilingual Matters, Clevedon, 1988, pp. 442–455.

[47] R. Levitan, A. Gravano, J. Hirschberg, Entrainment on backchannel-preceding cues, in: Proceedings of ACL, 2011.

[48] R. Levitan, J. Hirschberg, Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions, in: Proceedings of interspeech, 2011.

[49] R. Levitan, A. Gravano, L. Willson, S. Benus, J. Hirschberg, A. Nenkova, Acoustic-prosodic entrainment and social behavior, in: Proceedings of HLT/NAACL, 2012, pp. 11–19.

[50] J.D. Matarazzo, A.N. Weins, R.G. Matarazzo, G. Saslow, Speeech and silence behaviour in clinical psychotherapy and its laboratory correlates, in: J. Schlier, H. Hunt, J.D. Matarazzo, C. Savage (Eds.), Research in Psychotherapy, 3, American Psychological Association, Washington, DC, 1968, pp. 347–394.

[51] L.L. Namy, L.C. Nygaard, D. Sauerteig, Gender differences in vocal accommodation: the role of perception, J. Lang. Soc. Psychol. 21 (2002) 422–432.

[52] C. Nass, Y. Moon, B.J. Fogg, B. Reeves, D.C. Dryer, Can computer personalities be human personalities?, Int J. Hum–Comput Stud. 43 (2) (1995) 223–239.

[53] M. Natale, Convergence of mean vocal intensity in dyadic communication as a function of social desirability, J. Pers. Soc. Psychol. 32 (5) (1975) 790–804.

[54] K. Niederhoffer, J. Pennebaker, Linguistic style matching in social interaction, J. Lang. Soc. Psychol. 21 (4) (2002) 337–360.

[55] Nenkova, A. Gravano, J. Hirschberg, High frequency word entrainment in spoken dialogue, in: Proceedings of ACL/HLT, 2008, pp. 169–172.

[56] M.L. Newman, J.W. Pennebaker, D.S. Berry, J.M. Richards, Lying words: predicting deception from linguistic styles, Pers. Soc. Psychol. Bull. 29 (2003) 665–675.

[57] J.W. Pennebaker, R.J. Booth, M.E. Francis, Linguistic Inquiry and Word Count: LIWC (2007).

[58] A. Pentland, To signal is human, American Sci. 98 (2010) 204–210.

[59] J.S. Pardo, On phonetic convergence during conversational interaction, J. Acoust. Soc. Am. 119 (4) (2006) 2382–2393.

[60] M.J. Pickering, S. Garrod, Toward a mechanistic psychology of dialogue, Behav. Brain Sci. 27 (2004) 169–226.

[61] I. Poggi, F. D'Errico, Dominance signals in debates, in: A.A. Salah et al. (Eds.), HBU 2010, LNCS 6219, Springer-Verlag, Berlin Heidelberg, 2010, pp. 163–174.

[62] W. Putnam, R. Street, The conception and perception of noncontent speech performance: implications for speech accommodation theory, Int. J. Sociol. Lang. 46 (1984) 97–114.

[63] R. Ranganath, D. Jurafsky, D. McFarland, It's not You, it's Me: Detecting flirting and its misperception in speed-dates, in: Proceedings of EMNLP, 2009.

[64] D. Reitter, J.D. Moore, Predicting success in dialogue, in: Proceedings of ACL, 2007.

[65] M.J. Richardson, K.L. Marsh, R.W. Isenhower, J.R.L. Goodman, R.C. Schmidt, Rocking together: dynamics of intentional and unintentional interpersonal coordination, Hum. Movement Sci. 26 (2007) 867–891.

[66] J.B. Sexton, R.L. Helmreich, Analyzing cockpit communications: the links between language, performance, and workload, Hum. Perform. Extreme Environ. 5 (2000) 63–68.

[67] J. Sherblom, C. La Riviere, Speech accommodation and the effects of cognitive uncertainty and physiological arousal upon it, in: Proceedings of the Annual Meeting of the Speech Communication Association, 1987.

[68] S.L. Shullman, The illusion of devil's advocacy: how the justices of the Supreme Court foreshadow their decisions during oral argument, J. Appellate Pract. Process 6 (2) (2004).

[69] E. Shriberg, To ''Errrr'' is human: ecology and acoustics of speech disfluencies, J. Int. Phonet. Assoc. 31 (1) (2001) 153–169.

[70] R.A. Simmons, P.C. Gordon, D.L. Chambless, Pronouns in marital interaction, Psychol. Sci. 16 (2005) 932–936.

[71] A. Stenström, Pauses in monologue and dialogue, in: J. Svartvik (Ed.), London-Lund Corpus of Spoken English: Description and Research, Lund University Press, Lund, 1990.

[72] S. Stoyanchev, A. Stent, Lexical and syntactic priming and their impact in deployed spoken dialogue systems, in: Proceedings of NAACL, 2009.

[73] O.W. Stewart, M. Corley, Hesitation disfluencies in spontaneous speech: the meaning of um, Lang. Linguist. Compass 4 (2008) 589–602.

[74] M. Swerts, Conversational fillers as markers of discourse structure, J. Pragmat. 30 (1998) 485–496.

[75] H. Traunmüller, Analytical expressions for the tonotopic sensory scale, J. Acoust. Soc. Am. 88 (1) (1990) 97–100.

[76] M.E. Van den Berg, Language planning and language se in Taiwan: social identity, language accommodation, and language choice behavior, Int. J. Sociol. Lang. 1986 (59) (1986) 97–116.

[77] Ward, D. Litman, Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora, in: Proceedings of the SLaTE Workshop on Speech and Language Technology in Education, 2007.

[78] N.G. Ward, S.K. Mamidipally, Factors affecting speaking-rate adaptation in task-oriented dialogs, in: Proceedings of 4th Speech Prosody, 2008.

[79] J. Welkowitz, R.N. Bond, S. Feldstein, Gender and conversational time patterns as Japanese–American adults and children in same- and mixed-gender dyads, J. Lang. Soc. Psychol. 3 (1984) 127–138.

[80] M. Yaeger-Dror, The influence of changing vitality on convergence toward a dominant linguistic norm: an Israeli example, Lang. Commun. 8 (1988) 285–306.

[81] J. Ytsma, Bilingual classroom interaction in Friesland, in: A. Holman, E. Hansen, J. Gimbel, J.N. Jorgensen (Eds.), Bilingualism and the Individual, Multilingual Matters, Clevedon, 1988, pp. 53–68.