Entrainment in spontaneous speech: the case of filled pauses in Supreme Court hearings

Štefan Beňuš*, Rivka Levitan** and Julia Hirschberg**
* Constantine the Philosopher University in Nitra and Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
** Columbia University, New York, USA
sbenus@ukf.sk, {rivka, julia}@cs.columbia.edu

Abstract— **The aim of this paper is to contribute to our understanding of communicative social signals through a study of entrainment in the quality of filled pauses produced by Justices and lawyers during oral arguments of the Supreme Court. We report a tendency that the similarity between immediately adjacent filled pauses between a Justice and a lawyer correlates with the favorability of the Justice's vote. In addition to learning more about our cognitive abilities, we believe that better understanding of accommodation patterns is useful for increasing number of applications relaying on spoken interactions between humans and computers.**

I. INTRODUCTION

Reference [1] argues that "social species are likely to develop honest signals, a reliable communication system that serves to coordinate behavior between individuals." These communicative social signals called *honest signals* by [1] include primarily the patterns in timing, energy and intonational variability present in the interaction. This is because these signals seem to be evolutionarily pre-dating language and might thus be orthogonal, or even masked by, the semantic content of our utterances.

If the coordinating and accommodation among interlocutors happens primarily at the level of honest signals, the intervals of high accommodation should correlate with regions of low semantic content and thus low information density. Filled pauses such as *uh* or *um* seem ideal candidates for studying these communicative social signals since they are rather frequent in spontaneous conversations, they have very low information load, their prosodic features are profoundly redundant, they are subject to accommodation to the interlocutor [2], and their timing serves many functions especially in the system of turn-taking; e.g. [3].

This paper addresses three areas of research: filled pauses in spontaneous speech, entrainment (or accommodation) between interlocutors, and the communicative dynamics in the judicial domain. All three areas have been extensively studied in the past. Rather than treated as imperfection of speech, recent research has found that filled pauses play important and interesting roles in human communication. For example, they heighten the attention of the listener and help retaining the information following filled pauses in memory [4], they signal discourse and intonational structure [5], inform the listener about meta-cognitive states and conversational problems of the speaker [6].

Entrainment (also sometimes called accommodation, convergence, adaptation, etc.) is a pervasive feature of human interactions in which the behavior of one person becomes influenced by, and eventually more similar to, the behavior of his/her conversational partner. In speech, entrainment has been observed at multiple linguistic and paralinguistic levels; see e.g. [8] for a review of extensive literature.

Finally, the dynamics between the participants in court proceedings has also been extensively studied. Most relevant for this paper are findings that Justices and lawyers during Supreme Court oral arguments coordinate in terms of various function word usage, and that this coordination reflects the power-relationship (lawyers coordinate to Justices more than vice-versa and more so to unfavourable Justices, i.e. those that end up voting against them) [9]. This and other findings, however, rely on the textual features extracted from the transcripts while studies of acoustic and prosodic features of speech are scarce in this domain.

The aim of this paper is to contribute to our understanding of communicative social signals through a study of entrainment in the quality of filled pauses produced by Justices and lawyers during oral arguments of the Supreme Court. In addition to learning more about our cognitive abilities through investigating spoken accommodation, better understanding of accommodation patterns is useful for an increasing number of applications relaying on spoken interactions between humans and computers that are being developed. Several studies have shown that humans' ease and naturalness of accommodation is not limited to our conversations with other people but readily extends to human-machine conversation, e.g. [10], [11]. It is plausible that communications with machines will be more effective and perceived as more pleasing if the machine has the ability to accommodate to some of the features of the speech of the human who interacts with the machine. This is probably because humans perceive accommodation as a naturally human-like feature and perform it effortlessly, albeit unconsciously, when they interact with other humans. In this sense, the ability of machines to accommodate to the speech of the humans they interact with might lead to the perception of machines as more natural. Conversely, it has also been shown that humans may consciously decrease the similarity (i.e. dis-entrain) to people in order to increase their social distance to the interlocutor (e.g. [12]) or to show a negative attitude toward the interlocutor [13]. It seems that efficient human-machine communication systems should be sensitive to

978-1-4673-5188-1/12/$31.00 ©2012 IEEE

(dis)entrainment both in the recognition and synthesis domains. In this sense, our paper contributes to research that aims to merge cognitive capabilities of various levels.

II. METHODOLOGY

A. Corpus

Data for this paper come from the recordings of one full term of hearings in 2001 conducted by the US Supreme Court. The recordings were manually transcribed, the transcripts were checked for authenticity and filled pauses and other non-lexical information were added by professional transcribers. Finally, the words from the transcripts were semi-manually aligned to the available single-channel speech signal [14].

Analyzed data consist of 76 oral arguments, each of which usually takes little less than an hour. The petitioner and respondent sides are represented typically by one or two lawyers who present their cases and are questioned by the justices within an allowed time limit. In this term, nine justices include the Chief Justice Rehnquist (REHN) and Justices Breyer (BREY), Ginsburg (GINS), Kennedy (KENN), O'Connor (CONN), Scalia (SCAL), Souter (SOUT), Stevens (STEV), and Thomas (THOM). There were 198 lawyers appearing in these arguments but since some lawyers argued more than one case, the corpus contains speech from 150 lawyers.

In addition to the audio signal and the transcripts, the corpus also includes the identification of turns for each speaker and crude identification of speaker overlaps. Moreover, we also used the Spaeth database [15], which is a publicly available resource that includes a wealth of coded information about the cases and judges' votes. In particular, we used the information on weather a justice voted for or against the petitioner or respondent and since we knew which lawyer represented which case, we could determine if a justice "voted in favor of" or "against" a particular lawyer.

B. Material and feature extraction

The corpus contains multiple words or word fragments associated with a filled pause marker. However, four of these words are by far the most common: *ah*, *uh*, *eh*, and *um*. Their frequencies are shown in Table 1 and they represent 99.9% of all filled pause tokens and 2.29% of all words in the corpus. This rate is comparable to the filled pause rates in other corpora; e.g. [16].

TABLE I.
FREQUENCIES AND DISTRIBUTIONS OF FOUR MAJOR FILLED PAUSE TYPES IN THE CORPUS

Filled pause	Count	FP-Frequency	Word-frequency
uh	11935	67.0	1.53
um	2529	14.2	0.32
ah	1744	9.7	0.22
eh	1598	9.0	0.20
misc.	91	0.01	0.0
Total	17897	100	2.29

To answer the question if and how the similarity between the quality of filled pauses uttered by lawyers and those uttered by justices varies with the justice votes, we first extracted the word and FP counts for each speaker and docket. To test the effect of entrainment in FP type, we calculated a relatively naïve measure of entrainment for each justice-lawyer pair following [17] adapted here as (1).

$$entr(w) = |(count_{S1}(w)/ALL_{S1} - count_{S2}(w)/ALL_{S2})| \quad (1)$$

Hence, this measure reflects (an absolute value of) the difference between the fraction of times a particular word *w* – in our case the type of filled pause – is used by two speakers S_1 and S_2 – in our case a lawyer-Justice pair. As we discussed in section II-A above, publicly available data from Supreme Court hearings can be used to determine whether a justice voted in favor or against a particular lawyer. Using this information, we simply divided all Justice-lawyer pairs into two groups (in-favor vs. against, factor DECISION) and tested the effect of this factor on the distribution of dependent variable ENTR(FP).

The distribution of filled pause types (*uh*, *ah*, *eh*, and *um*) represents only a course and discrete information on the quality of vowels in filled pauses. Furthermore, since the database transcripts were most likely produced by multiple coders, this information about perceived vowel quality might be somewhat subjective. We do not have information about the inter-labeler agreement in coding filled pause types.

To obtain continuous and more objective information on the vowel quality of filled pauses, we extracted the values of the first and second formant (F1 and F2 respectively) from the midpoint of each filled pause. These values were then normalized to the perceptually meaningful Bark scale [18].

Two tests were conducted with these values. First, we determined the centroids (means) for the filled pause distributions in the F1-F2 space for each speaker and docket if the speaker produced at least three filled pauses in the docket. We then calculated the Euclidean distance between these centroids for each lawyer/justice pair within a docket, which represented the dependent variable for the second test. Hence, if the Euclidean distance is smaller in the $Justice_x$-$Lawyer_y$ pair than in the $Justice_x$-$Lawyer_z$ pair, we may say that $Lawyer_y$ is more similar to $Justice_x$ than $Lawyer_z$ in terms of vowels in filled pauses. This first test thus assesses the global similarity in the quality of filled pauses between a Justice and a lawyer over the entire duration of the oral argument and call the dependent variable calculated in this way ENTR(GLOBAL).

In the second test we aimed at assessing the similarity in the filled pauses on a more local basis. The oral arguments frequently include multiple miniature one-on-one dialogues between a lawyer and a Justice containing several turn exchanges. The global measure of entrainment would not detect if a lawyer adjusts her speech to the justice to whom she currently responds. One rather crude way of exploring this kind of entrainment is to calculate the distance between justice-lawyer filled pauses but rather than using the means over the entire session as above, use only adjacent filled pauses for the justice-lawyer pair. Hence, our dependent variable in this test is populated by Euclidean distance values for each

pair of adjacent filled pauses between a lawyer and a Justice. This is illustrated in Fig. 1 below in which each line represents one value of our dependent variable ENTR(LOCAL) corresponding to the Euclidean distance between the F1 and F2 values of the two filled pauses at the two ends of the line segment.

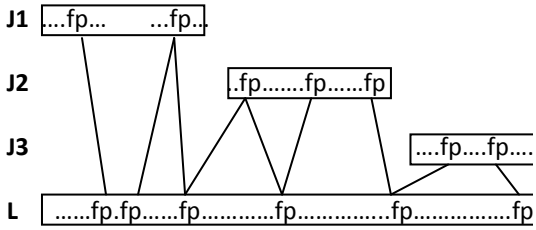

Figure 1. Idealized speech (represented as boxes) of justices (J1, J2, J3,…) and a Lawyer (L) with interspersed filled pauses (fp)

In both of the tests described above, the dependent values for justice-lawyer pairs were tested for the effect of factor DECISION described above.

III. RESULTS

A. Descriptive observations from the justices: Rates

We first look at the distribution of the four major filled pause types in the speech of the nine justices. Fig. 2 shows the frequencies for each type as well as when they are pooled for each justice.

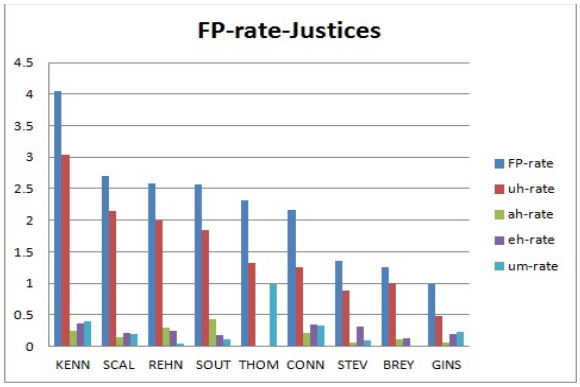

Figure 2 Distribution of frequencies of four major filled pause types and pooled together for individual Justices

There is one rather surprising observation already previewed by the data in Table I: Filled pauses are primarily produced as *uh* and not as *um*. The reason for surprise is that according to [7], *um* signals rather more thorough planning problems while *uh* tends to signal lexical access problems. Given the complexity of legal discourse and familiarity of both Justices and lawyers with legal terminology and jargon, one would expect a more even ratio between the two major types of filled pauses.

The data from Justice Thomas can be safely disregarded since they represent only seven filled pauses in the entire corpus whereas the other eight justices produce between 308 (STEV) and 1802 (SCAL) filled pauses. It is well known that Justice Thomas speaks very little during oral arguments and the number of filled pauses strongly correlates with the amount of speech.

If [7] is right, we would expect that lawyers should have a greater ratio of *um* than Justices since they are questioned and need to provide credible answers "on the spot". Table II shows the rates of filled pause types separately for Justices and lawyers. We see that, indeed, the greatest difference is in *um*-rate that is higher for lawyers than for Justices. 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 results was obtained from a chi-square test with 3 categories ({*uh, ah*, *eh}*, *um,* and *other*). Hence lawyers use the nasalized filled pauses more frequently than Justices.

TABLE II.
RATES OF THE MAJOR FILLED PAUSE TYPES SEPARATELY FOR JUSTICES AND LAWYERS

Filled pause	FP-rate Justices	FP-rate Lawyers
uh	1.59	1.49
um	0.16	0.42
ah	0.18	0.25
eh	0.23	0.19
Total	2.17	2.37

B. Descriptive observations from the justices:Quality

Fig. 3 shows the vowel quality of all filled pauses longer than 40ms produced by eight Justices in the F1-F2 space and the Bark scale. Despite great overlaps, it can be seen that the transcription of three non-nasalized varieties roughly corresponds to their quality. Filled pauses transcribed as *eh* (blue) are in general more front and high than *uh* for all speakers whereas those transcribed as *ah* (green) are somewhat lower for some speakers. This corresponds to the differences between canonical vowels represented by these transcriptions.

The second observation is 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.

C. Testing the relationship between similarity in filled pause production and votes

In section II-B we described three dependent variables for assessing the relationship between the similarity in filled pause use in justice-lawyer pairs and whether the decision by a justice in the pair was favorable to the lawyer in the pair. The first dependent variable – ENTR(FP) – measured the difference in filled pause rates per docket in justice-lawyer pairs. A t-test with DECISION (favor vs. non-favor) as a factor showed that similarity in filled pause rates between justices and lawyers had no significant effect on whether the justice votes were favorable to the lawyers ($t = 0.24$, $df = 1401$, $p = 0.81$). Investigating if individual Justices conform to the overall result, separate t-tests were run. There was no significant result and the t-values were split evenly (4 to5) in terms of polarity; positive t-values show that similar fp-rates correspond to favorable votes. Therefore, similarity in the frequency of filled pause use between a justice and a lawyer had no

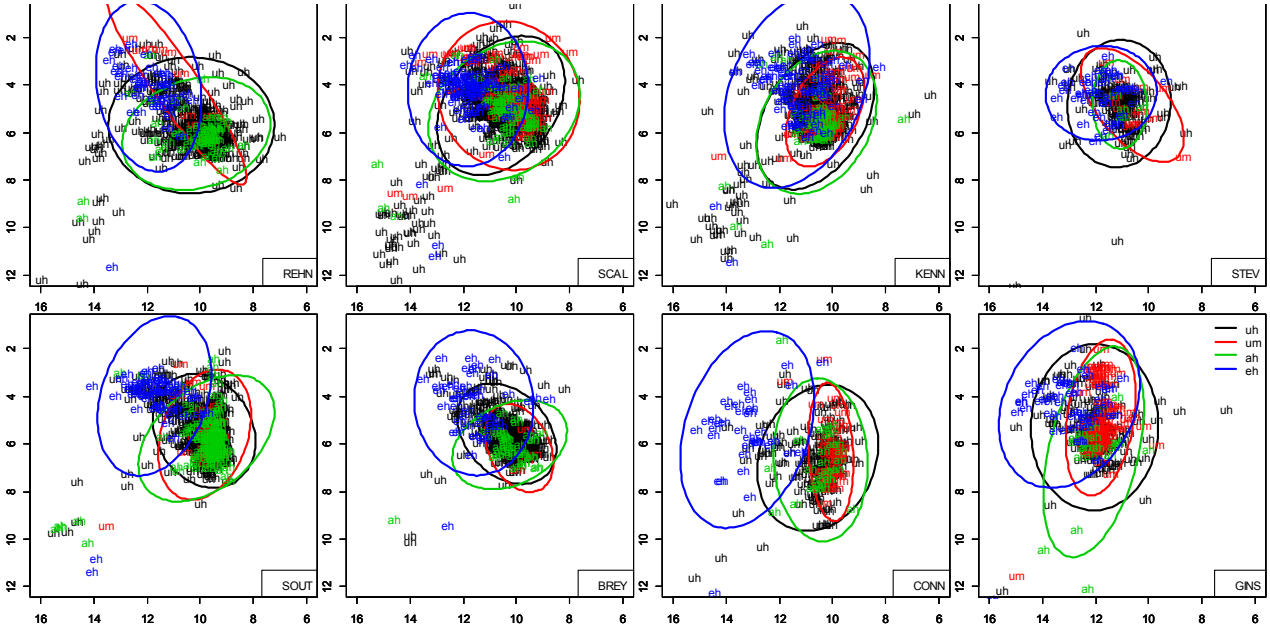

Figure 2. First two formants in Bark scale of all filled pauses produced by 8 Justice with ellipses showing 95 percent confidence intervals

significant effect on the direction to which the justice leaned in his or her decision.

The second dependent variable – ENTR(GLOBAL) – tested if the similarity in the quality of filled pauses had any effect on the direction of justice votes. This analysis showed that the mean distances for justice-lawyer pairs in which the justice "voted for" the lawyer, were smaller than the distances in the pairs where the justice voted "against" the lawyer. However, this effect was very small and non-significant ($t = 1.25$, $df = 1081.5$, p-value = 0.21). We again tested for individual Justices separately and obtained no significant result at alpha = 0.05 and only one result approaching significance at alpha = 0.1. In this test, six of the eight Justices had more similar filled pauses to the lawyers they voted for including the result approaching significance at $p < 0.1$, and two Justices produced filled pauses more similar to the lawyers they voted against. We reiterate that none of these tests were significant at $p < 0.05$. Hence, the results from global entrainment in the quality of filled pause corroborate the result from filled pause rates that there is no significant relationship between filled pause use and the direction of the justice votes.

Finally, the third dependent variable – ENTR(LOCAL) – tested the similarity between adjacent filled pauses in justice-lawyer turns and whether this (dis)similarity correlated with justices' votes. This local measure of entrainment did show a significant effect of factor DECISION: mean distance between adjacent filled pauses was smaller in those justice-lawyer pairs in which the justice gave a favorable vote than in those pairs in which the justice gave a non-favorable vote ($t = 2.26$, $df = 982.1$, $p = 0.024$). A similar result was obtained when we tested separately for lawyers representing the petitioner side ($t = 1.98$, $df = 432.2$, $p = 0.049$). Even in separate tests for individual justices, one test showed significance ($t = 2.13$, $p = 0.035$) and one tendency ($t = 1.75$, $p = 0.084$) in the same direction as the overall test. In these either separate tests, seven t-values were positive (including the two already mentioned above) and only one was non-significantly negative. Therefore, the local measure of entrainment between justices and lawyers in terms of the quality of filled pauses produced implies that if a lawyer produces filled pauses similar to the justice s/he currently talks to during the oral argument, there is a tendency that this justice might vote favorably in that lawyer's case. We discuss in section IV why this preliminary finding should be taken with caution.

D. Initial observations from turn-taking behavior

The use of filled pauses is closely linked to turn-taking behavior. This is because one of the primary functions of filled pauses is to signal the interlocutor that the speaker wishes to hold the floor. Occasionally, filled pauses might be used also to signal the wish to grab the floor, to relinquish it [3], or to acknowledge the need for information [19]. In section II-A, we mentioned that the corpus contains crude identification of overlaps between speakers. Using this information we calculated the interruption rate for each justice-lawyer pair. Interruption here is defined in such a way that $Speaker_1$ spoke before an overlap region, $Speaker_1$ and $Speaker_2$ both spoke during the overlap region, and $Speaker_2$ spoke after the overlap region. Note, however, that a small number of these interruptions might also correspond to so called cooperative overlaps that function to support, affirm, or acknowledge what the other speaker is saying [20]. Having this measure of interruption rate, we then tested its correlation with the favorability of the votes. We found a very weak positive tendency that interrupting (i.e. when lawyers interrupt Justices) correlates with favorable decisions for lawyers; $r = 0.14$, $p = 0.053$. However, it has been also observed that the more a justice talks to or questions the lawyer, the less likely s/he is to vote for the lawyer [21], [22], [23]. Given the fact that longer mini-dialogues between justices and lawyers probably correlate with the number of interruptions, our preliminary finding requires a thorough follow up investigation but points to the area of turn-taking as a promising domain for mining

for more information about the dynamics of Supreme Court hearings.

IV. DISCUSSION AND CONCLUSIONS

The goal of this paper was to investigate the potential of filled pauses as a communicative social signal of entrainment and attitudes between interlocutors. The judicial domain was represented by the corpus of speech produced by Justices and lawyers during oral arguments of the US Supreme Court. We tested three measures of entrainment and if they correlate with favorability of the Justices' votes. Two of these tests – naïve measure of similarity in filled pause rates and global similarity in vowel quality for the entire argument – showed no relationship. The third measure tapped into more local entrainment by measuring the pair-wise similarity between adjacent filled pauses in Justice-lawyer pairs. This local measure of entrainment was affected by justices' favorability toward the lawyer's case: mean difference between adjacent filled pauses was smaller in those Justice-lawyer pairs in which the Justice gave a favorable vote than in those pairs in which the Justice gave a non-favorable vote.

There are several reasons for taking this finding with caution. First, this is a first preliminary investigation applying most rudimentary measures and analyses. Second, filled pause similarity might be related to phonetic context for which we did not control. For example, if one speaker tends to produce filled pauses following *the*, their quality will be influence by this preceding vowel. So, the entrainment between speakers might be in terms of other features than the phonetic ones. Third, the directionality in adjacent filled pauses FP might have an effect. Our third measure was bi-directional; hence, a filled pause from a lawyer could have been counted for a preceding Justice as well as for the following one. Finally, Justices base their decisions on many other aspects besides the oral arguments and lawyers know about their leaning before the oral argument [9]. Thus, any finding from the speech of oral argument is necessarily only partial.

We plan to expand this research by testing the prosodic features of the filled pauses primarily focusing on pitch and intensity as well as voice quality features.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under Grant No. IIS-0803148 and in part 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] A. Pentland, "To Signal Is Human," *American Scientist*, 98, pp. 204-210, 2010.

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

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

[4] O. W. Stewart, M. Corley, "Hesitation disfluencies in spontaneous speech: the meaning of um," *Language and Linguistics Compass* 4, pp. 589–602, 2008.

[5] M. Swerts, "Conversational fillers as markers of discourse structure, " *Journal of Pragmatics* 30, pp. 485–496, 1998.

[6] S. E. Brennan, M. Williams, Maurice, "The feeling of another's knowing: prosody and conversational fillers as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language* 34, pp. 383–398, 1995.

[7] H. H. Clark, J. E. Fox Tree, "Using uh and um in spontaneous speaking, " *Cognition* 84, pp. 73–111, 2002.

[8] J. Hirschberg, "Speaking More Like You: Entrainment in Conversational Speech", in Proc. INTERSPEECH, 2011.

[9] C. Danescu-Niculescu-Mizil, L. Lee, B. Pang, J. Kleinberg, " Echoes of power: language effects and power differences in social interaction," Proceedings of the 21st international conference on World Wide Web, pp. 699-708, 2012.

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

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

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

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

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

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

[16] E. Shriberg, "To "Errrr" is Human: Ecology and Acoustics of Speech Disfluencies," *Journal of the International Phonetic Association* 31(1), pp. 153-169, 2001.

[17] A. Nenkova, A., A. Gravano, and J. Hirschberg, "High frequency word entrainment in spoken dialogue," Proceedings of ACL/HLT 2008 (short paper), pp. 169-172, 2008.

[18] H. Traunmüller, "Analytical expressions for the tonotopic sensory scale," *Journal of the Acoustical Society of America*, *88*(1), pp. 97–100, 1990.

[19] Š. Beňuš, A. Gravano, J. Hirschberg, "Pragmatic aspects of temporal accommodation in turn-taking," *Journal of Pragmatics*, 43(12), pp. 3001-3027, 2011.

[20] D. Tannen, *You just don't understand: Women and men in conversation*, New York: Ballantine, 1998.

[21] John G. Roberts Jr., "Oral Advocacy and the Re-emergence of a Supreme Court Bar," *Journal of Supreme Court History.* 30 (1), 68-81, 2005.

[22] Sarah Levien Shullman, "The Illusion of Devil's Advocacy: How the Justices of the Supreme Court Foreshadow Their Decisions During Oral Argument." *Journal of Appellate Practice and Process.* 6 (2), 2004.

[23] Lee Epstein, William Landes, and Richard 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.