Visual Filler: 視覚刺激提示による伝送遅延状況下での円滑な遠隔対話の実現

Visual Filler: Visual Stimuli to Facilitate Smooth Communication over TV Conference System with Transmission Delay

西川 猛司 † 川嶋宏彰† **松山 隆司** †
Takeshi Nishikawa Hiroaki Kawashima Takashi Matsuyama

1. 伝送遅延状況下での話者交替

円滑な会話において，話者交替の自然さは，会話に参加する複数の話者間での間合いの予測によって支えられている．しかし，遠隔対話においては，伝送遅延によって発話が相手話者に伝達されるタイミングが遅れるため，話者の間合いが強制的に冗長となり，自然な話者交替がしばしば困難となる．

ここで，対面会話においては，音声発話だけでなく身体動作（視線や口元の動き）が，発話権の解放や取得時刻を表現・推定する際に用いられていると考えられる [1]．このとき，身体動作を視覚刺激として解釈すると，視覚刺激が話者の主観的な間合いに影響を与えると考えられる．そこで本研究では，伝送遅延の生じる遠隔対話において，会話支援システムが会話画面中に視覚刺激を生成することによって，話者の主観的な間合いを補間し，その冗長さを解消する方法を提案する．

この視覚刺激は，間合いを補間する点においては，音声発話におけるフィラー（「えーと」など）に対応しているため，この視覚刺激を Visual Filler とよぶことにする．間合いの補間のためには，音声フィラーを先行話者に提示する方法も考えられるが，この方法では言語的内容を伝達する上での弊害になりかねない．一方，Visual Filler はあくまで身体動作の代用であり，音声に干渉せず，補助的に間合いの補間が可能である．

次節では，まず実際の対話について予備的分析を行ない，話者交替の間合いに対して身体動作が与える影響を考察する．この考察に基づき，3 節以降で Visual Filler による間合いの補間について述べる．

2. 身体動作が間合いに与える影響の分析

発話の代わりに身体動作を用いることでも，聞き手が感じる間合いに影響を与えるであろうか．話者交替における，身体動作と間合いの関係を調べるにあたり，まず対話分析を行なった．話者交替の間合いは，対話の種類や話者の個性に大きく依存し，しばしば対話の状況設定の決め方が問題となる．このひとつの解決法として，分析対象に演芸の対話を選択した．なぜなら，演芸においては，会話を観客に心地よく聞かせるために間合いが最適に調整されており，さらに，対話の状況が一般的な自由会話に比べて限定できると考えられるためである．

本稿では，会話内容が日常会話に近いものに設定されている演芸として，一人二役によって，複数話者の会話を表現する落語を分析した．落語演者によって同時に演じられる役柄はただ一人であるため，先行話者である役柄と後続話者である役柄の発話はオーバーラップできない．このとき，落語演者が役柄の交替を表現するために，特徴的な頭部動作（左右の動き）を用いていることに着目し，演者が，発話ではなく頭部動作のタイミング制御によって話者交替を表現し，主観的な間合いを調整しているという仮説を立てた．そこで，先行話者である役柄の発話終了時刻に対する，役柄交替を表現するための頭部動作の開始時刻までの時区間を解析した．さらに，二者の演者による会話演芸である漫才における，先行話者（ここではツッコミ役）の発話終了時刻に対する，後続話者（ボケ役）の発話開始時刻までの時区間を解析することで比較を行った．この結果を表 1 に示す．

表 1: 落語における先行発話終了時刻に対する頭部動作の開始タイミングと漫才における先行発話終了時刻に対する後続発話の開始タイミングの時区間の比較（単位 [msec]）．各分布の四分位点と平均を示す．

	1st.Qu.	Median	3rd.Qu.	Mean
Rakugo	-58	43	113	42
Manzai	-105	34	203	64

表 1 より，2 つの分布は，中央値の差が 10msec，平均の差が 20msec 程度とほぼ一致する．さらに，第一四分位と第三四分位の差も 100msec に満たない．したがって，これらの分布は類似しているといえる．この結果より，主観的な間合いを調整する上で，落語における頭部の動作が，二者間会話における発話と同様の働きを持つ可能性が考えられる．これは，身体動作が主観的な話者交替の間合いに影響を与えることを示唆する．これらの演芸の詳細な比較は稿を改めて報告するものとし，次節以降では，身体動作の代わりに，人工的に視覚刺激を生成する方法である Visual Filler について検討する．

3. 提案手法: Visual Filler

話者交替における発話権の移動方法について，従来よりさまざまな議論があるが [2, 3, 4]，本稿においては，発話権の解放と取得は，具体的な行為と密接な関係があるものの，あくまで，話者各自の主観の中で生じるものと解釈する [5]．すなわち，発話権の解放意志や取得意志は，さまざまなモダリティの信号（音声発話，身体動作）によって，意図的，あるいは無意識的に表現され，これらの信号から，各話者は相手話者の意志を推定する．このとき，話者は各自，主観的に，発話権の解放から取得

†京都大学情報学研究科, Grad. Sch. of Informatics, Kyoto Univ.

までの間合い（主観的な話者交替の間合い）を感じることになる．

3.1 話者交替における主観的な間合い

まず，話者間の意思伝達に遅延が存在しない，通常の会話における話者交替について，本稿での仮定を述べる．

発話権の解放と取得は，発話という行為と密接に関連すると考えられる．先行話者は，重要な内容をすべて発信した後においては，発話中であっても話者交替を許す．したがって，発話権解放は先行発話中に生じると考えるのが自然である．一方，後続話者は，自身の発話開始をもって相手に話者交替を表明するため，発話権取得は，後続発話と同時であると考えるのが自然である．

このとき，先行話者と後続話者は，それぞれの主観の中で話者交替の間合いを感じるとする．先行話者による発話権の解放時刻を $T(R^p)$，後続話者による発話権の取得時刻を $T(A^s)$ と表し，先行話者が推定・認識する後続話者の発話権取得時刻を $T(W_A^p)$，後続話者が推定・認識する先行話者の発話権解放時刻を $T(W_R^s)$ と表す．さらに，時刻 $T(X)$ に対する時刻 $T(Y)$ までの時区間（およびその長さ）を $I(X,Y)$ と表記すると，先行話者の主観的な間合いは，$I(R^p, W_A^p)$，後続話者の主観的な間合いは，$I(W_R^s, A^s)$ となる．

各話者の主観的な間合い $I(R^p, W_A^p)$ と $I(W_R^s, A^s)$ には，話者にとって自然に感じられる区間長が存在する．本稿では，その区間長は，話者間で共有されていると仮定し，N で表現する．ここで，単純化のために以下の場合のみを考える．まず後続話者は，主観的な間合いを N に一致させるように $T(A^s)$ を調整する．一方で先行話者は，自分が発話権を解放した時刻 $T(R^p)$ に対して，相手話者が適切なタイミング（N だけ経過した時点）で取得することを期待している．これを発話権取得期待時刻 $T(E_A^p)$ とよぶことにすれば，$T(E_A^p) = T(R^p) + N$ である．発話権の解放意志および取得意志が，後続話者および先行話者にそれぞれ正しく推定された場合，$T(A^s)$ は $T(E_A^p)$ に一致し，先行話者の主観的な間合い $I(R^p, W_A^p)$ の大きさは N に一致する（図 1）．これは，先行話者が話者交替を自然に感じることを意味する．

一方，実際の会話においては，先行話者の発話権解放時刻を後続話者が誤って推定する場合がある．このとき，発話権の取得時刻 $T(A^s)$ は，発話権取得期待時刻 $T(E_A^p)$ と一致せず，$I(R^p, W_A^p)$ が N と一致しない状況が生じる（後続話者の発話権取得時刻を先行話者が誤って推定した場合も同様）．ここで，この間合いが N から離れるほど，話者が感じる不自然さは上昇すると考えられる．

3.2 伝送遅延による間合いの冗長さ

遠隔対話においては，映像と音声の伝送遅延が存在するため，意志の伝達に遅延が生じる．本稿では，問題の単純化のため，遅延は双方向かつ一定であると仮定し，遅延時間を一定値 D で表す．

通常の会話においては，発話権の解放意志と取得意志のいずれかの推定が失敗したときに，両者の主観的な間

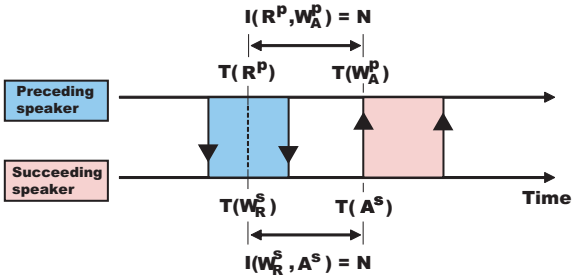


図 1: 通常の会話における話者交替の間合い

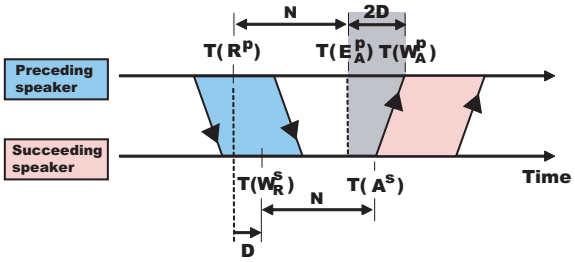


図 2: 遅延状況における間合い（$2D$ だけ冗長となる）

合いが一致しない状況が生じたが，遅延状況においては，これらの推定が正しく行なわれると仮定しても，伝送遅延によって，発話権の解放意志や取得意志の推定が意志の表現時刻よりも遅れる（図 2）．そのため，先行話者の主観的な間合いの大きさ $I(R^p, W_A^p)$ は式 (1) を満たす．

$$\begin{aligned} I(R^p, W_A^p) &= I(R^p, E_A^p) + I(E_A^p, W_A^p) \\ &= N + 2D \end{aligned} \quad (1)$$

ここで，時区間 $I(E_A^p, W_A^p)$ において，先行話者は，いずれの話者も発話権取得意志を持たないと思い不安になり，しばしば再発話を行おうとする．一方，後続話者は，主観的な間合い $I(W_R^s, A^s)$ を N と一致させながら発話を行う．その結果，先行話者の再発話と，後続話者の発話にぶつかりが生じ，円滑な会話を行うことが困難となる．したがって，$I(E_A^p, W_A^p)$ が冗長な間合いであることを先行話者に意識させないようにすれば，両者に話者交替を自然に感じさせることができ，さらに発話のぶつかり頻度を減少できると期待される．

3.3 Visual Filler による間合いの補間

本稿では，遠隔対話において生じる間合いの冗長さを解消するために，支援システムによって，間合いを補間するための視覚刺激 (Visual Filler) を話者に提示する方法を提案する．

支援システムは，遠隔対話において画面を介して会話が進むことを利用し，先行話者が見ている画面に対して視覚刺激を挿入する（図 3）．ここで，この提示タイミングが重要であり，本稿では，区間 $I(E_A^p, W_A^p)$ 内に視覚刺激を提示するものとする．この提示によって，$I(E_A^p, W_A^p)$ を冗長な区間として先行話者に意識させないことが可能になると考えられる．

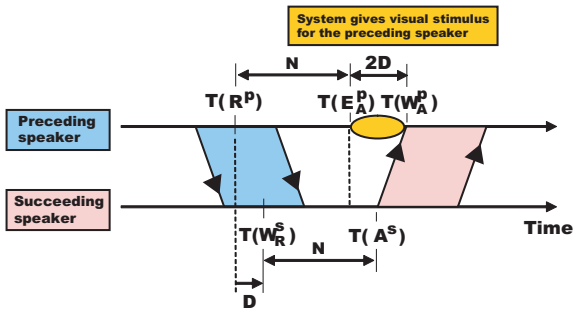


図 3: Visual Filler による冗長な間合いの補間

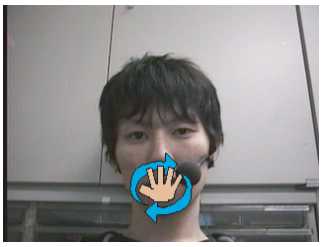
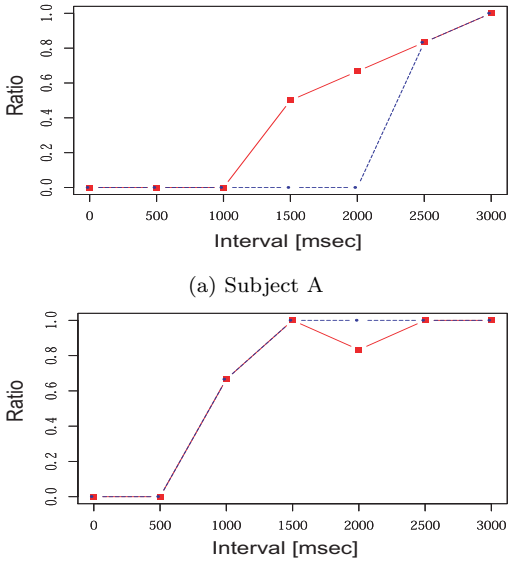


図 4: Visual Filler の挿入（被験者が見る相手話者の映像に対して回転動作するアイコンを重畳して表示）



(a) Subject A

(b) Subject E

図 5: 恒常法による評価結果（5 人中 2 人の結果を示す）．横軸は発話移行区間長，縦軸は被験者が「早く返答してほしい」と評価した割合を表す．四角および丸（それぞれ実線および破線でつないでいる）は，Visual Filler の提示の無，有の条件に対する値をそれぞれ表す．ただし，Visual Filler の提示は 1500msec 以上の条件について行なっているため，1000msec 以下の条件の数値は，提示の有無にかかわらず同じ値である．

4. 実験

4.1 手続き

本実験では，会話状況として二者間会話における一度の話者交替を想定した．被験者は，椅子に座った状態でディスプレイに表示された相手話者に発話を行なうものとし，その被験者の発話に対して，相手話者が返答を行なうものとする．このとき，被験者の発話終了時刻に対する相手話者の発話開始時刻（発話移行区間長）について，被験者に話者交替の不自然さを評価させた．このとき，相手話者の映像に対して Visual Filler を挿入し，挿入しない場合との被験者の評価を比較した．

Visual Filler に用いる視覚刺激は，間合いの冗長さを話者に意識させないことを要求される．そこで，回転動作するアイコンを映像に挿入することによって，被験者の意識をアイコンに引きつけることをねらった（図 4）．

本実験では，評価方法として，映像や音声の刺激閾調査に用いられる測定法の一種である，恒常法を用いた．恒常法は，複数の刺激を用意し，被験者への一刺激の提示と刺激に対する評価を，ランダムかつ十分な数だけ反復する方法である．本実験では，被験者と相手話者との一度の話者交替を反復し，反復のたびに発話移行区間長を変動させ，被験者に不自然さを評価させた．発話移行区間長は，0，500，1000，1500，2000，2500，3000msec の 7 段階とした．さらに，1500msec 以上の 4 つの段階については，Visual Filler を挿入する場合を条件に加えた．なお，Visual Filler の挿入する時間は，被験者の発話終了時刻から 1000msec 経過した時点から，相手話者の発話開始時刻までとした．

以上の 11 条件について，各条件が 6 回ずつランダムに提示されるように，話者交替を計 66 回反復した．各反復において，被験者は「もっと早くに返答してほしいと思ったか否か」についての二者択一により評価した．なお，同じ条件の試行を反復するために，相手話者の映像には，あらかじめ撮影した動画を用いた．

4.2 条件

被験者は 5 人であり，映像中の相手話者と親しい人間とした．これは，被験者と相手話者が初対面であると，違和感が生じると共に，間合いの共有という仮定が崩れる恐れがあるからである．

被験者と相手話者の発話内容を以下に示す．

- 被験者: 休暇中に行ったことの報告
- 映像中の相手話者: 「へぇ，そうなんだ」

発話内容によっては，発話移行区間長の変化により，相手話者の発話の意味や目的が，被験者に異なって理解される可能性がある（例えば「それはいいね」などの肯定，否定がある場合）．そのため本実験では，相手話者の発話への印象が，発話移行区間長の変化にそれほど依存しないと考えられる，単に報告を受理するような中立的な発話内容を選択した．

4.3 結果

被験者各自について，相手話者の発話に対して早く返答してほしいと思った割合を算出したところ，5 人中 4 人の被験者では，Visual Filler の提示によって，早く返答してほしいと思った割合は低下した．その例を図 5(a) に示す．ここで，Visual Filler 提示の有無による早く返

表 2: 映像フィラーの挿入による不自然さの割合の低下度．低下度は Visual Filler を提示しない条件での不自然さから提示した条件での不自然さを減じた値である．

	Interval [msec]			
	1500	2000	2500	3000
Subject A	0.50	0.67	0	0
Subject B	0.33	0.50	0	0
Subject C	0	0	0.83	0.50
Subject D	0.33	0.33	0	0
Subject E	0	-0.17	0	0

答してほしいと思った割合の差を，便宜上，低下度とよび，不自然さを表す値として用いる．表 2 は，被験者ごとの，1500msec 以上の条件における低下度を表す．

表 2 より，被験者 A，B，D の 3 人については，1500msec と 2000msec の 2 つの条件について，不自然さに低下がみられる．とくに，1500msec の条件において，不自然さは 0 に低下した．一方，2500msec 以上の条件においては，低下度は 0 であり，不自然さは 1 に近い値を示した．また，被験者 C については，2000msec 以下の条件において，Visual Filler 提示の有無によらず，不自然さは 0 であった．一方，2500msec 以上の条件において，不自然さは増加したが，Visual Filler の提示によって，0.5 ポイント以上低下した．

一方，被験者 E のみ Visual Filler の提示によって不自然さに低下がみられない．ここで，被験者 E の評価結果を図 5(b) に示す．図より，被験者 E の不自然さは，1000msec 以上の条件において 0 から上昇している．本実験においては，被験者の先行発話終了時刻から 1000msec の時点から Visual Filler を被験者に提示したため，この結果は，被験者 E が Visual Filler を提示開始時刻において，すでに不自然さを感じていることを表している．

5. 考察

5.1 Visual Filler の効果

4. 節の実験では，5 人中 4 人の被験者で，Visual Filler の提示によって話者交替に対する不自然さが低下する傾向がみられた．これは，Visual Filler が話者の意識を引きつけ，発話の遅れによる冗長な間合いを被験者に意識させないことに成功したためであると考えられる．

5.2 効果の持続時間

5 人中 3 人の被験者では，発話移行区間長が 1500msec と 2000msec の条件において不自然さが低下したが，2500msec 以上の条件においては不自然さが低下しなかった．これは，Visual Filler が被験者に意識させない間合いには限度があり，発話の遅れが大きいとき，Visual Filler は被験者の意識の引きつけを継続できないことを示唆している．ひとつの解決策としては，1sec 程度ごとに動きや刺激を切り替えるなど，Visual Filler を適切に設計することが考えられる．

5.3 Visual Filler の生成タイミング

5 人中 1 人の被験者については Visual Filler による不自然さの低下がみられなかった．これは，この被験者は Visual Filler の生成時刻 (1000msec) においてすでに不自然さを感じていたことに原因があると考えられる．このように，不自然さを感じる間合いについては個人差が大きく，Visual Filler の提示タイミングを適応的に決定する方法を検討する必要がある．さらに，話者が話者交替を不自然さを感じるタイミングは，話者交替が生じる文脈に大きく依存すると考えられる．

5.4 Visual Filler の種類

アイコンの回転動作を突発的に開始するのではなく，常時，点滅や回転をさせておき，点滅や回転のスピードをそれぞれ速めるような Visual Filler についても，低下度に基づいて定量的に評価を行ったが，その結果，突発的な開始，点滅スピードの変化，回転スピードの変化の順に，不自然さの低下度は大きかった．このように，どのような種類の Visual Filler を用いるかも重要といえる．

6. おわりに

本稿では，遠隔対話における話者交替の間合いの冗長さを補間するために，視覚刺激 (Visual Filler) を話者に対して提示する方法を提案した．被験者実験より，回転動作などの動きを持つアイコンを Visual Filler として被験者に提示したときに，被験者が感じる，相手話者の発話の遅れに対する不自然さが低下することが明らかになった．一方で，Visual Filler の効果の持続時間には限度があること，提示タイミングによっては効果がみられないことが明らかになった．

今後は，心理実験の試行回数を増やすと共に，実際の会話状況で十分な評価を行う必要がある．さらに，Visual Filler の種類として，頭部や口元の動きの映像生成を含む様々な視覚刺激を設計することや，提示タイミングの決定方法として，文脈や個人に適応させる方法を検討する必要がある．また，実用上は発話開始や終了時の検出が問題となるが，口元の動きと音声の同期関係に基づく発話区間同定法を現在検討中である．

謝辞: 本研究の一部は，科学研究費補助金 18049046 の補助を受けて行った．

参考文献

[1] Kendon, A., Some function of gaze-direction in social interaction, Acta Psychologica, 26, 22–63, 1967.

[2] Duncan, S. J., and Fiske, D. W., Face-to-Face Interaction, Lawrence Erlbaum, chapter 11, 1977.

[3] Nagaoka, C., Komori, M., Nakamura, T. and Draguna, M. R., Effects of Receptive Listning on The Congruence of Speaker's Response Latencies in Dialogues, Psychological Reports, 97, 265–274, 2005.

[4] Shriberg, E. E., Spontaneous Speech: How People Really Talk, and Why Engineers Should Care. Proc. Eurospeech, 1781–1784, 2005.

[5] Sacks, H., Schegloff, E. A., and Jefferson, G., A simplest systematics for the organization of turn-taking for conversation. Language, 50, 696–735, 1974.