

会話における話者のうなずきと 発話音声のプロミネンスの時間関係

天谷 晴香 (東京大学大学院総合文化研究科) †

Timing Relationships between Prominences of Speaker Head Nods and Pitch Movements

Haruka Amatani (The University of Tokyo)

要旨

発話音声のプロミネンスと発話に伴うジェスチャーのストローク・ピークは一致することが多いと言われる。McNeill(1992)はこれを *phonological synchrony rule* によるものとした。それらの厳密な時間関係を調査した研究のひとつに Nobe(1996)がある。Nobe は英語話者の類像ジェスチャーのストローク・ピークが発話音声のピッチ・ピークと同期または先行するとしてビート・ジェスチャーもまたストローク・ピークを音声のピッチ・ピークと同期または先行させる。日本語話者の頭部ジェスチャーのうなずきには、ビート・ジェスチャーと似たふるまいを見せるものがあるが、発話のピッチ・ピークとうなずきのストローク・ピークは同期または固定した先行関係が成立しているか。アクセント語と無アクセント語を分類した上で、うなずきとピッチの各ピークの時間関係を明らかにする。

1. はじめに

話者は発話時、言語情報だけではなく非言語情報を豊富に発している。文字や音声情報に加えて、ジェスチャーなどの身体動作情報を加えたマルチモーダルな会話研究は、より包括的な記述で、会話の全体像を捉えようとするものである。

発話に伴う身体動作は、視線の動きや頭の動き、手によるジェスチャーなどがある。特に頭部動作のうなずきは日本語話者に特徴的に多く見られる動きである。メイナード(1993)によると、アメリカ英語話者の約3倍、日本語話者は会話中にうなずいている。

発話そのものに加え、うなずきや動作などが協調して会話のリズムを作っているという分析を、Erickson and Schultz(1982)は英語会話について行った。ザトラウスキー(1997)は、日本語会話のリズムは英語会話のそれとは質的に違うが、日本語会話でも非言語情報が会話リズムに貢献する可能性を示唆した。

発話音声の強弱や上昇下降調とジェスチャーの強弱や方向が一致するとしたのが Bolinger(1983)である。Bolinger のこの主張で、方向が一致するとした部分は後に否定されている(Loehr 2004)。

しかし、音声のピッチの上昇位置にジェスチャーが発現する現象は実際に見られる(Cave et al. 1996 他)。McNeill(1992)はこのような音声とジェスチャーの *phonological synchrony rule*(音韻共時法則)と呼んだ。Nobe(1996)は、表象ジェスチャーが英語話者によって発せられる時、そのジェスチャーの主要部分であるストロークのピークが音声のピッチ・ピークと同時かもしくは先行して起こると報告している。これと同様に、ビート・ジェスチャーのストローク・ピークが音声のピッチ・ピークと同時か先行して起こった(Loehr)。

発話に伴ううなずきと音声のピッチの関係はどのようになっているだろうか。日本語の単語には語彙アクセントがある。語彙アクセントのない言語の研究から、音声のピッチ・ピークとジェスチャーのストローク・ピークが同期しやすいことが言われている。語彙アクセントは急激なピッチ下降を生じさせ、音声的に際立っている。語彙アクセントによるピッチの動きは語彙アクセントによらないピッチの動きよりうなずきと同期しやすくある

† amatani.haruka@gmail.com

かどうか、会話音声とうなずき頭部動作を詳細に分析することで、明らかにしたい。

2. 発話に伴ううなずき

うなずきと言うと、聞き手のあいづちとしてのうなずき動作がまず思い起こされるが、話し手も発話しながらうなずき動作を相当数行っている。

メイナード(1993)は日本語話者の会話において、話し手のうなずきと聞き手のうなずきが同程度の数、出現したことを報告している。また、庵原ら(2004)は話し手のうなずきが聞き手のうなずきより多く出現したことを報告している。

3. うなずきの種類と出現位置

メイナードは話し手のうなずきの役割に、「同意」「承認」「強調」「節のマーカ」「肯定」「リズム取り」「ターンの受け継ぎに関係する機能」があるとした。

また、前田ら(2003)は、話し手のうなずきは聞き手の反応を要求するものとしたが、金田(2007)は「対人的な機能は発話全体から見出されるものであり」、顎刻み(話し手のうなずき)が有するものではないとしている。

金田は、話し手のうなずきの出現位置として、「発話末(句末・文末)」および「重要な箇所最初のモーラ」を挙げている。

重要な箇所の最初のモーラに身体動作が現れるという現象は、話し手のうなずきを視聴覚韻律(audiovisual prosody)として研究されてきた身体動作のひとつとして考える根拠となる。

視聴覚韻律には、話し手のふるまいを観察した研究から、音声のピッチの動きに付随する、フランス語話者の眉の動き(Cave et al. 1996)、英語話者と日本語話者の頭部動作(Yehia et al. 2002)などが挙げられる。また Yehia et al.が出した頭部動作と音声のピッチ動作は一致しやすいという結果から Munhall et al. (2004)は日本語のデータを使って 3D アニメーションの頭部映像を作り、知覚実験を行って、頭部動作を付随させた音声は聞き取りやすくなるという結果を報告している。Krahmer & Swerts (2007)は手のビート・うなずき・眉の動きを、オランダ語の音声的強調の置かれる単語に付随させて、発話したものを、視聴者に見せる知覚実験を行っている。動きが付随した場合、強調がより強く感じられたとしている。

4. 動作としてのうなずきの分析

細馬・富田(2011)は、ジェスチャー区間の観点から聞き手のうなずきを2種類に分類した。ジェスチャー区間は、Kendon(2004)が用いたジェスチャー単位の最も小さなレベルである。細馬・富田は Kendon や細馬(2008)の、主に手のジェスチャー分析で用いられてきたジェスチャー単位を援用し、頭部動作であるうなずきを分析している。

ジェスチャー単位は1つもしくは複数のジェスチャー句から形成される最も大きなレベルである。1つのジェスチャーが1つのジェスチャー句を成し、ジェスチャー句はジェスチャー区間から成り立っている。最も際立ったジェスチャー区間はストロークである。そして、ストロークの前の予備的な動きは準備区間、ストロークの後の元の定位置に戻る動きは復帰区間とされる。また、これらの区間の間に、保持と呼ばれる動きの止まる区間が存在しうる。

金田(2007)は、話し手の「うなずき」は聞き手のうなずきと異なり、顎を正面の位置から上げてからその後下げる「リズムを刻む時のような顎の動きである」ため、「顎刻み」と呼ぶとした。細馬・富田は、この金田の分析にジェスチャー区間を適用して、話し手のうなずきはPS型と分析している。

このように、うなずきを複数のジェスチャー区間に分けて分析することで、非常に細かい時間単位においてうなずきの生起位置を特定することができる。

5. 会話データと分析

5.1 データ

会話は実験室で録音・録画された。2人の参加者によるものである。それぞれ話者 A,B とする。対面に向かい合った状態で会話しており、ビデオカメラは2台で、部屋の隅からそれぞれの身体全体を一人ずつ画面におさめている。音声はマイクをヘッドセットで装着し録音した。

参加者は実験の始めに約10分の別々のアニメーションを視聴した。30分の会話の中で、互いに自分の視聴した内容について説明し合った。分析に用いたデータは30分の会話の内、最初の10分である。内容は主に互いが観たアニメーションについて説明し合ったものだった。

5.2 音声の分析

会話音声のアノテーションには、X-JToBI(Maekawa et al. 2002)を用いた。また、分析に使用したソフトウェアはPraat(Boersma and Weenink 2014)である。ピッチの動きとして、基本周波数(F0)の動きを採用し、記述した。

本研究では、トーンとアクセント句の判定を重点的に行い、分析対象とした。トーンのアノテーションから、語彙アクセント位置や語彙アクセントに伴わないF0の下降位置を抽出した。また、うなずきの共起を判断する範囲を、アクセント句とした。アクセント句は多くの場合、文節に対応する範囲である。以下で「アクセント語・無アクセント語に伴ううなずき」と言った場合、その語を含むアクセント句内にうなずきのピーク位置があることを意味する。

5.3 動作の分析

動作アノテーションには、細馬・富田が聞き手のうなずきについて行ったように、Kendonのジェスチャー単位を援用した。分析に使用したソフトウェアは、ELAN(Sloetjes and Wittenburg 2008)である。ビデオをコマ送りで視聴し、頭部が動き出すタイミングや軌道が変わるタイミングを記述した。

金田の指摘のように、話者のうなずきが「顎刻み」である場合、むしろその動作のストロークは上方向の動きである可能性が考えられる。上方向の動きのピーク位置、すなわち顔・顎が一番高い位置にある点と、下方向の動きのピーク位置、すなわち顔・顎が一番低い位置にある点が、どちらも可能なうなずきのストローク・ピークになりうる。本研究では、下方向の動きを主に分析対象にしている。ただし、下方向の動きの開始位置で顔・顎は最も高い位置にあることになるため、その時点を仮に上方向の動きのストローク・ピーク位置として、分析に用いた。そのことを明記して、以下、分析結果を報告する。

また、話し手・聞き手のうなずきの区別については、聞き手のあいづちに発声のある場合、それと同時に発せられるうなずきは発話に伴ううなずきとした。

6. 分析結果

6.1 うなずきと語彙アクセント

まず語彙アクセントにうなずきに伴いやすいかを調べるため、アクセント語を含むアクセント句とアクセント語を含まないアクセント句においてうなずきの出現率に差があるか測った。表1、表2にそれぞれ、話者A、話者Bの結果を示した。

表1. アクセント句の語彙アクセントの有無とうなずきの有無の関係 (話者A)

	うなずき有り	うなずき無し
語彙アクセント有り	164	219
語彙アクセント無し	44	78

表2. アクセント句の語彙アクセントの有無とうなずきの有無の関係 (話者 B)

	うなずき有り	うなずき無し
語彙アクセント有り	162	369
語彙アクセント無し	62	141

それぞれにカイ 2 乗検定を行った所、アクセント句の語彙アクセントの有無は、うなずきの生起率に影響していないことが分かった。アクセント語にも無アクセント語にも同様にうなずきが伴ったり伴わなかったりすることが分かった。

6.2 うなずきのストローク・ピークとピッチ・ピーク

話者 A、B それぞれに、うなずきのストローク・ピークであると考えられる下方向の頭部動作の最も低い時点と、音声のピッチ・ピークが最も高い時点の関係を以下、図に示す。また、同時に、話し手のうなずきのストロークが上方向の頭部動作である可能性をふまえて、上方向の頭部動作の最も高い時点と、音声のピッチ・ピークの時間関係も図に示す。

図 1～図 4 が話者 A、図 5～図 8 が話者 B の観測結果である。それぞれ、図 1・2 と図 5・6 がアクセント語に共起したうなずきの起きた回数を表しており、図 3・4 と図 7・8 が無アクセント語に共起したうなずきの個数を表している。また、図 2・4 と図 6・8 が下方向のうなずきのピークとピッチ・ピークの時間差を示しており、図 1・3 と図 5・7 が上方向のうなずき (たりえる頭部動作) とピッチ・ピークの時間差を示している。

グラフの X 軸の単位は「秒」である。この「秒」によって、動作ピークとピッチ・ピークの時間差が表されており、その差は動作ピークの起きた時間から、ピッチ・ピークの起きた時間を引くことで算出された。

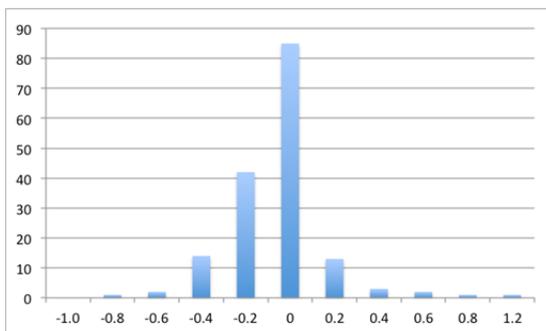


図 1. 語彙アクセントと上方頭部動作のピークの差 (話者 A)

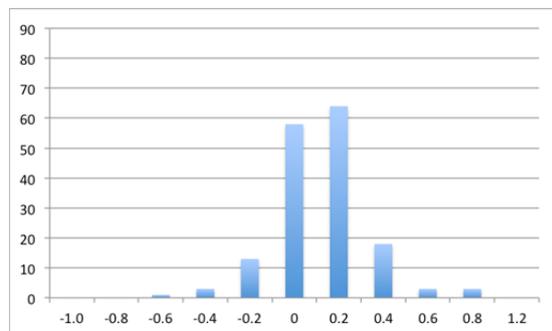


図 2. 語彙アクセントと下方頭部動作のピークの差 (話者 A)

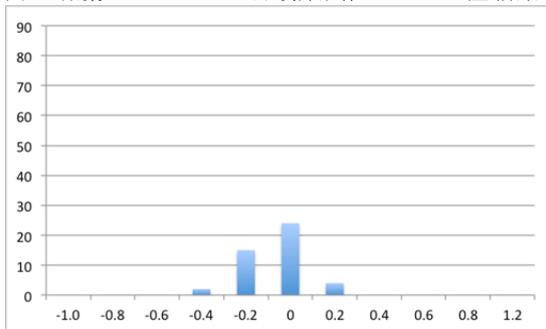


図 3. 無アクセントと上方頭部動作のピークの差 (話者 A)

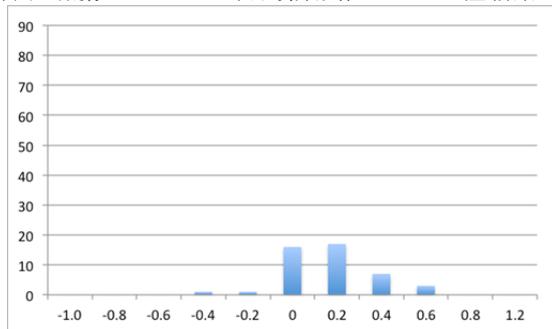


図 4. 無アクセントと下方頭部動作のピークの差 (話者 A)

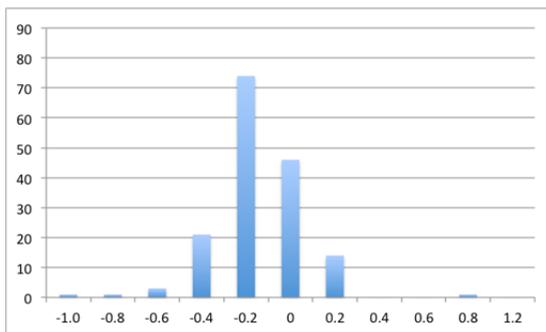


図 5. 語彙アクセントと上方頭部動作のピークの差 (話者 B)

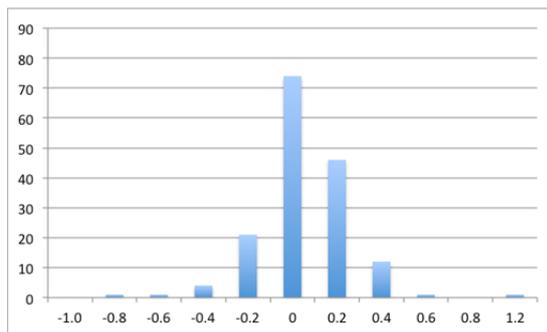


図 6. 語彙アクセントと下方頭部動作のピークの差 (話者 B)

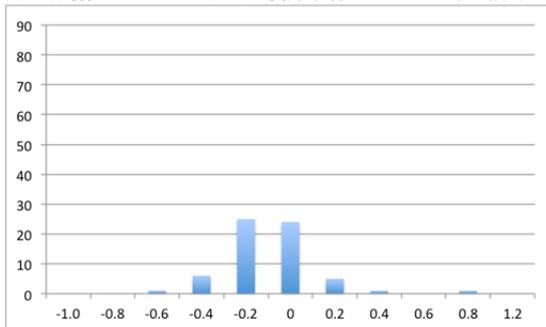


図 7. 無アクセントと上方頭部動作のピークの差 (話者 B)

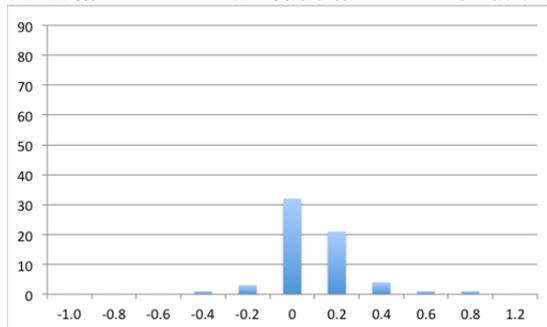


図 8. 無アクセントと下方頭部動作のピークの差 (話者 B)

話者 A において、上方向の頭部動作のピークは語彙アクセントの有無に関わらず、ピッチ・ピークから 0 秒～0.2 秒の間に最も多かった。また、下方向の頭部動作のピークは語彙アクセントの有無に関わらず、ピッチ・ピークから 0.2 秒～0.4 秒の間に最も多く見られた。

話者 B においては、上方向の頭部動作のピークは語彙アクセントの有無に関わらず、ピッチ・ピークから -0.2 秒～0 秒の間に最も多かった。下方向の頭部動作のピークは語彙アクセントの有無に関わらず、ピッチ・ピークから 0 秒～0.2 秒の間に最も多く見られた。

これらの結果から、うなずきの発生する音声のピッチ・ピークに対するタイミングは、語彙アクセントの有無より個人差が影響する可能性が示唆される。また、個人差はあるが、うなずきはピッチ・ピークとかなり近い位置で起こっていることも分かった。

話者のうなずきを上方向、下方向どちらの動作と捉えるかについては、結果から話者 A では上方向、話者 B では下方向と言えそうな結果になっている。しかし、バラツキもあるため、個々のうなずきを観察し判定するのが望ましく、話者のうなずきの型をひとつに決定することは難しい。

7. おわりに

うなずきのストローク・ピークと音声のピッチ・ピークは、非常に近接して起こることが、詳細な動作と音声の分析からわかった。ただし、語彙アクセントの有無はうなずきの発生率に影響していなかった。

音声とジェスチャーのリンクを言語的な要素に基づくものでなく、運動のメカニズムから説明しようとするのが、Rusiewicz (2012)である。言語産出の過程でなく、運動実行の過程を音声とジェスチャーは共有しており、そのために各々のプロミネンスが共起するとする。音声とジェスチャーの運動実行過程の共有を言語産出モデルに取り入れたものに、Tuite (1993)がある。

音声とジェスチャーは協調して発話リズムを作っていると考えられる。そのリズムがどこまで言語的制約に依拠し、どこから運動的なリズムによって説明され得るものかについて示唆を得られるよう、今後、頭部動作と音声の構造を詳細に分析していきたい。

謝 辞

本研究で分析に用いた会話データを収録し、筆者に使用を許可して下さった University of Victoria 博士課程の Thomas Magnuson 氏に感謝いたします。

文 献

- Boersma, P. and Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4, retrieved 4 October 2014 from <http://www.praat.org/>
- Cave, C., Guitella, I., Bertrand, R., Santi, S., Harlay, F., and Espesser, R. (1996). About the relationship between eyebrow movements and f0 variations. In H.T. Bunnell and W. Isardi (eds.), *Proceedings of the 4th International Conference on Spoken Language Processing*, pp. 2175-2178.
- 庵原彩子、堀内靖雄、西田昌史、市川嘉(2004)「自然対話におけるうなずきの機能に関する考察」電子情報通信学会技術研究報告. HCS, ヒューマンコミュニケーション基礎 104(445), 13-18.
- 金田純平(2007)「発話中の話者による頭の動き-のけぞりと顎刻み-」国際シンポジウム「日本語『音声言語』の教育と基礎資料」神戸大学、2007年12月
- Krahmer, E. and Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57, 396-414.
- Loehr, D.P. (2004). *Gesture and intonation*. Doctoral dissertation, Georgetown University.
- 前田真季子、堀内靖雄、市川嘉(2003)「自然対話におけるジェスチャーの相互関係の分析」情報処理学会研究報告. HI, ヒューマンインターフェース研究会報告 9, 39-46.
- Maekawa, K., Kikuchi, H., Igarashi, Y., and Venditti, J. (2002). X-JToBI: an Extended JToBI for spontaneous speech. In *INTERSPEECH*.
- メイナード 泉子(1993)『会話分析』くろしお出版
- McNeill, D. (1992). *Hand and Mind*. University of Chicago Press.
- Munhall, K.G., Jones, J.A., Callan, D.E., Kuratate, T., and Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, 15-2, 133-137.
- Nobe, S. (1996). *Representational gestures, cognitive rhythms, and acoustic aspects of speech: A network/threshold model of gesture production*. Doctoral Dissertation, University of Chicago.
- Sloetjes, H. and Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*. (LREC 2008).
- Rusiewicz, H.L. (2012). Synchronization of prosodic stress and gesture: a dynamic systems perspective. *Gesture and Speech in Interaction*.
- ザトラウスキー ポリー(1997)「日本語の談話のリズム分析 『息の合った』会話を例に」 p.101-148, 茂呂雄二(編)、『対話と知』新曜社
- Yehia, H.C., Kuratate, T. and Varikiotis-Bateson, E. (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*, 30, 555-568.