

発話-行動インタラクションにおける相互適応と韻律情報の影響

小松孝徳[†] 鈴木健太郎[†] 植田一博[†]
開一夫^{†,††} 岡夏樹^{†††}

私たちは日常生活において、共通の言語体系をもたないにもかかわらず、発話を通じて親密なコミュニケーションをしているような関係を観察することができる。筆者らは、お互いに相手に適切な反応をすることを継続的に学習していく相互適応と、発話中の言語情報に依存しない韻律情報とがこのようなコミュニケーションにおいて重要な役割を果たしていると仮定し、これを検証するための実験を行った。実験の結果、うまくコミュニケーションを成立できた被験者組において相互適応が観察され、発話中の韻律情報であるピッチ成分が相手に対して注意を喚起する働きをしていたということが明らかにされ、共通の言語体系を持たない者同士においても、相互適応と発話中の韻律情報を利用することで、親密なコミュニケーションを成立できるということが確認された。

Effects of Mutual Adaptation and Prosodic Information in a Speech/Action Interaction

TAKANORI KOMATSU,[†] KENTARO SUZUKI,[†] KAZUHIRO UEDA,[†]
KAZUO HIRAKI^{†,††} and NATSUKI OKA^{†††}

In our daily life, we could observe the intimate relationship between interactants who do not share the same language. In this paper, we assume that both mutual adaptation, in which interactants learn to respond appropriately to one another's behavior, and the prosodic information in a speech play a significant role in establishing this kind of communication. To study these effects, we prepared a game experiment. The result showed that the successful pairs showed mutual adaptation, and that pitch information as a prosodic information in speeches had a consistent role of drawing a partner's attention. Therefore, this study identified that these effects had an important role in establishing smooth speech/action interaction, even in a restricted communication.

1. はじめに

私たちは日常生活において、共通の言語を持たないにもかかわらず親密なコミュニケーションをしているような関係を目にする事が出来る。例えば、ペットと飼い主の関係もそのような関係の一つであろう。飼い主がペットと何らかのインタラクションを取ろうとする時、ペットが人間のコトバを理解できないことを知りながら、彼らはペットに話しかけたり、人間のコトバで指示を出したりする。最初、ペットは飼い主の言っていることを理解できないが、飼い主が話しかけ方を変化させたり、それに応じてペットも反応を変

化させるといった両者の試行錯誤的な行動の変化が観察されるようになる。そして、お互いがお互いのことを学習していく経時的なインタラクションを繰り返すことで、最終的にペットは飼い主の発話を理解できるようになると考えられる(例えば、「お手」「伏せ」等)。筆者らはこのようにお互いのことを繰り返し学習していくプロセスを相互適応と呼ぶ。しかし、ペットは飼い主の用いている言語体系を獲得していないのに、どのようにして飼い主のコトバを理解しているのだろうか? この時ペットは、発話情報のうち文字で表現できる音韻情報から言語的な規則に基づいて意味を獲得しているのではなく、与えられた発話のある種の信号として自身の行動と結び付けることによる「素朴な意味理解」をしていると考えることができる。ここで筆者らは、発話情報のうち文字で表現することのできないような韻律情報がこのような素朴な意味理解において重要な役割を果たしていると考え、このような共通の言語体系を持たない二者においても、「相互適応」と

[†] 東京大学大学院 総合文化研究科
Graduate School of Arts and Sciences, The University
of Tokyo

^{††} 科学技術振興事業団
Presto, JST

^{†††} 松下技研株式会社
Matsushita Research Institute Tokyo, Inc.,

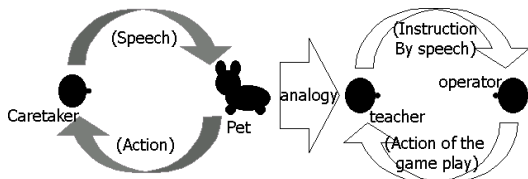


図 1 Simple interaction between a pet and its caretaker and between a teacher and an operator

「発話中の韻律情報」を利用することでスムーズなコミュニケーションを構築することができると仮定した。

本研究では、この仮定を検証するための簡単なゲーム実験を行った。この実験には、二人の被験者（一人はゲームの教示者、もう一人は操作者）が参加し、操作者は教示者から与えられる韻律情報のみによる音声教示に従って実際にゲームをプレイする一方、教示者は操作者の動きを観察しながら音声で指示を出す役割が与えられる。つまり、「スイカ割り」のような状況を再現しているといえる。このゲーム実験において被験者らが高得点を獲得できれば、共通の言語を持たない物同士に見られるようなコミュニケーションを獲得できたと言える。この実験の目的は、相互適応・韻律情報がコミュニケーション構築において果たす役割を、被験者の行動・教示者の発話音声の分析から明らかにすることである。

ここで得られる結果は、あらかじめ言語的なルールやデータを埋め込むこと無く人間とコミュニケーションをとることのできるエージェントの開発・より自然な形のマンマシンインターフェイスの実現などに多くの示唆を与えるであろう。

2. 実験環境・手順

本研究における実験環境を図 2 に示す。二人一組の被験者はお互いのことを見ることのできない別々の部屋に配置され（教示者は部屋 A、操作者は部屋 B）、それぞれの前にはお互いが同じ画面を見ることができる TV ゲームが置かれている。実験で使用したゲームは、操作者がラケットを動かして落ちてくるボールを拾うことができれば 10 点が与えられ、失敗すると 10 点減点されるという、スカッシュのようなゲーム (Pong) である。操作者は実際にラケットを動かし、教示者は操作者に対してボールを当てるように音声で教示を与えるが、その教示音声はローパスフィルタ (LPF) を通して操作者のいる部屋 B のスピーカから流される（教示者は自分の LPF が掛った声をリアルタイムで聞くことはできない）。LPF は発話音声の中の情報のうち、文字で表現できるような音韻情報を削除

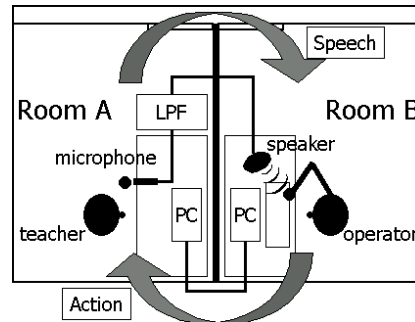


図 2 Game environment

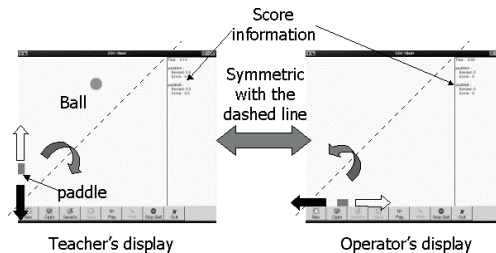


図 3 Different display setting

するが、韻律情報には影響を与えない。つまり、教示音声は LPF を通過することで、操作者は教示音声の韻律情報しか聞くことができなくなる。

本実験においては、図 3 のように教示者と操作者に異なった画面表示が割り当てられた。この状況において、教示者の画面では操作者にとってターゲットであるボールが表示されるが、操作者のそれでは表示されない。よって、操作者は自分のラケットと、ラケットがボールに当たったかどうかを示している得点の表示画面しか見ることが出来ないため、全ての状況を把握している教示者からの教示に従う必要がある。また、二者の画面は図 3 中の斜めの点線によって対称に折り返されている。その理由として、もし操作者が教示者の教示が「右」「左」のどちらかと想定し、教示者もその予測どおりの教示を与えた場合、操作者は LPF によって音韻情報を得ることができないにもかかわらず、リズムや発話の長さからでその教示を容易に推測できてしまうからである*。よって、このような予測を避けるために、このような画面設定が導入された。

実験の手順として、まず、実験者は被験者ペアに、「このゲーム実験の目的はお互いが協力して高い得点を取ることです」と説明する。次に、LPF の掛けられていない普通の音声とそれに LPF を掛けた音声を

* 例えば、「ボ、ボ」と聞こえると「ミギ」で、「ボ、ボ、ボ」と聞こえると「ヒダリ」と理解されるような状態。

表 1 Comparison of subjects' Hit Ratio

Category		Hit Ratio
Category 1		
Failed to understand any instructions	2pairs	0.29, 0.26
Category 2		
Succeeded in understanding the direction	4pairs	0.54, 0.51
		0.53, 0.55
Category 3		
Succeeded in understanding the distance	5pairs	0.52, 0.37
		0.78, 0.50, 0.63
*control	no instructions	(0.22, 0.21)
	without LPF	(0.68, 0.73)

比較のために聞かせる。この際、画面の設定がお互いに異なるということには言及しない。その後実験を開始し、10分プレイ2分休憩を1セットとして、2~3セットを行う。実験中の教示者と操作者の役割は固定で、実験の休憩中に顔を合わせて相談等をする機会とは与えられない。

3. 実験結果

11組の被験者組(22人, 20~28才, 日本人の男性:18人, 日本人の女性:4人, 被験者組においてお互いがすでに知り合いであることが条件)が実験に参加した。被験者11組のうち、2組が教示者のいかなる教示も理解できなかった。残り9組のうち4組が教示者の意図する方向にラケットを動かしたがボールにうまく当ることができなかった組で、残りの5組が教示者の意図する位置にラケットを動かしてボールにうまく当ることができた組であった。以降、前者4組を「方向理解を獲得出来た組」、後者5組を「距離理解を獲得できた組」と呼ぶ。統制群として、(1). 教示者から全く教示を与えられないグループ(教示無しグループ)、(2). 教示者からの教示にLPFが掛けられなかったグループ(LP無しグループ)の二種類のグループを用意した。Hit Ratioはボールをラケットに当ることができた「打率」を表したもので、教示を全く理解できなかった2組は教示無しグループとほぼ同じHit Ratioを示した一方、ほとんどのペアが二つの統制群の中間値を見せた。また、一部の理解度の高い組はLPF無しグループとほぼ同等のパフォーマンスを見せていた(表1参照)。図4は、教示理解度の典型的な遷移を表したグラフである。教示理解度とは、0点=操作者が教示者の意図する方向と逆に動いた場合、1点=教示者の意図する方向に動いたがボールに当られなかった場合、2点=教示者の意図する方向に動いて且つボールに当ることができた場合、のように各試

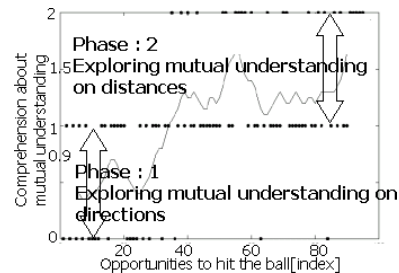


図 4 Transition of comprehension about instructions

行毎に点数を与えたものである。グラフ上の黒い点は教示理解度をプロットしたもので、曲線は10試行ごとの教示理解度の移動平均である。この図から、(1). 曲線が0-1の間を遷移しているときは方向理解を獲得しようとしているフェーズであり、1-2を遷移しているときは距離理解を獲得しようとしているフェーズである、(2). 方向理解は距離理解よりも先に達成される、という二つのことが理解できる。この方向・距離の二つの理解フェーズにおいて、教示音声の韻律情報(この実験の場合、特にピッチ情報)・操作者ラケットの動き・Hit Ratio・教示種類といった側面から実験の分析を行った。

3.1 フェーズ1. 方向理解フェーズ

11組中9組が方向理解を獲得することができた。このフェーズにおける理解プロセスはこれら9組においてほぼ共通であった。まず、教示者は操作者に対して、「端まで」「ちょっと」「逆だよ」等、様々な種類の教示を与えていた。この実験において非常に多く観察される教示として「上、上...」「下、下...」というものがあるが、これらのピッチ曲線を見ても(図5の左)、また実際の音声聞いても、この時点でこの両音声の区別をつけることは非常に困難であった。その後、教示者は徐々に教示の種類を減らし、最終的にほとんどの教示者は繰り返して「上、上、上...」「下、下、下...」と発話する二種類の教示に落ち着いた。この教示数の減少と同時に、これら二種類の教示のピッチ曲線・実際の音声ともに、区別がつきやすくなっていた(図5の右)。具体的には、前者の教示音声はLPFを介することで、一つの長い音として聞こえ、後者はぶつと途切れ途切れの音に聞こえる。操作者はこのような教示音声を聞くと、二種類の教示が与えられていることを理解し、同時にこれら与えられた音声と自分のラケットの行動とを結び付けることで教示の意味を理解しようとし始める。例えば、操作者が偶然にラケットでボールを当ることができた時、「今、与えられていた教示音声で、たった今自分のとった行動を示していた」ということが自発的に理解できる。このよう

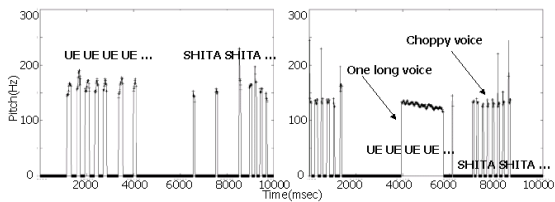


図 5 Pitch curves of 'UE' and 'SHITA' in Phase 1. left: 250 sec, right: 540 sec

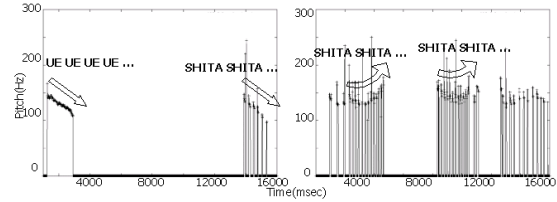


図 7 Pitch Curve of 'UE' and 'SHITA' in Phase 2. left: 1050 sec, right: 1270 sec

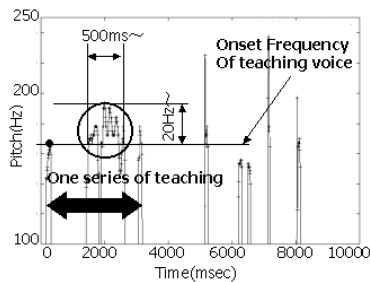


図 6 A high-pitched element in teaching voice

な試行錯誤的な理解の一方で、教示者は操作者に対して教示と行動との正しい対応を、教示音声の高ピッチ要素によって引き出すことに成功していた。具体的には、図 6 のように教示音声に約 20Hz 程発話開始時よりもピッチが高く、これが最低 0.5 秒継続する成分があると、操作者は直感的に教示者からの警告であると理解し、実際にこのような高ピッチの音声を見たときに、「この行動は間違っている...」と理解し、ラケットの動作方向をすぐに修正していた。よって、このフェーズにおいて、高ピッチ音声は操作者の現在の行動を否定する機能を果たしていた。

3.2 フェーズ 2. 距離理解フェーズ

この Pong ゲームにおいては、方向理解だけでは高得点は期待できない。なぜなら、どちらの方向にラケットを動かすのか (方向理解) というだけでなく、どの位置までラケットを動かすのか (距離理解) ということが、ラケットでボールを当てることにおいて非常に重要であるからである。方向理解を獲得できた 9 組の被験者組のうち、5 組が距離理解を獲得することができていた。これらの 5 組はおおまかに、STOP 教示を使用した組、しなかった組の二つに分けることができる。

3.2.1 グループ 1. STOP 教示使用

5 組中 2 組がこのグループに属する。「STOP」という音は、短く跳ねるような音であるので、操作者にとって文字通り「STOP」と理解することは非常に容易であり、これらの被験者は実験開始直後に「STOP」

教示の共通理解を獲得していた。これらの組の被験者は、繰り返しの「上, 上, 上...」(一つの長い音として聞こえる) を右方向、「下, 下, 下...」(ぶつぶつと途切れた音に聞こえる) を左方向と理解していた*。このような方法で、方向理解を獲得した場合、「上...」という教示の使用の際に、教示者はこれを単発で使用することができない。なぜなら、もし教示者が単発で「上」とだけ発話したならば、これは長い音ではないので、操作者は途切れ途切れ・ぶつぶつ切れる教示であると理解し、逆を意味する左方向に動いてしまうからである。それゆえ、右方向への教示は繰り返し発話によって行わなければならないため、目的地よりもオーバーランする傾向になっていた。これを解決するために、教示者は「上上..., STOP, 下下...」と教示することで、オーバーランした距離分を元に戻す教示を開始した。この教示の変化に対して操作者は、繰り返し「上」教示の直後に STOP が聞こえると、繰り返し「下」教示がくるということを学習し、実際に「上上...」の後に STOP 教示が与えられると、「下下...」と言われる前に逆の方向に動くようになった。つまり、操作者は教示者の「STOP」という発話を、文字通り「STOP」と解釈したり、「ちょっと戻れ」というように、解釈の仕方を状況に応じて使い分けていたのである。また同様に、教示者もこれに応じて STOP 教示の使用方法を使い分けていた。この一連の行動は、まさにこの実験系で観察された相互適応の一例であろう。

3.2.2 グループ 2. STOP 教示不使用

残りの 3 組がこのグループに属する。このグループの被験者も方向理解に関しては、前のグループと同様の理解の仕方をしていた。このグループにおいて特徴的だったのが、教示音声の発話開始時のピッチが高く、終端でピッチが低くなっており、下がり気味のイントネーションになっていたということであった (図 7 の左)。このような状況下において、操作者のラケッ

* 画面の角度が操作者と教示者で異なっているため、教示者にとっての上方向が、操作者にとっての右方向と一致している。

ト移動距離は比較的短いいため、長い距離の移動が必要な場合においてはボールに届かないといった動作が観察された。教示者は操作者に長い距離移動を達成させるため、発話終端のピッチの値を調整することでこの問題を解決していた。つまり、終端のピッチの値を上昇させることで長い距離移動が必要であることを示し、それを意図的に下げることで短い移動を表していた(図7の右)。このフェーズにおいて、発話の高ピッチ成分は、被験者の動きを急かす意味として解釈されていた。これはフェーズ1における高ピッチの理解と合わせて、操作者に対して広い意味で注意を与えていたと理解できる。

4. 考 察

4.1 韻律情報の与える影響

一連の実験解析から、高ピッチ音声は操作者の方向理解だけでなく距離理解を促進しているということが確認された。高ピッチ音声は未知教示に対する理解を促進しているということを確認するために、簡単な追加実験を行った。教示者があらかじめ理解しやすい「上」「下」教示を、操作者に対して(1). 高ピッチ無しで与える場合、(2). 高ピッチをつけて与える場合の二つの実験を行い、操作者が方向教示を理解するのに要した時間を比較した。その結果、高ピッチ音声を使用した場合の方が方向教示理解まで約120秒かかったのに対して、無い場合の方が約180秒であった。このことから、高ピッチ音声は操作者の未知教示理解の促進に影響を与えているということが確認できた。

このような高ピッチ音声などの韻律情報が発話の意味理解に与える影響について、いくつかの先行研究(マザリーズ音声、韻律ブートストラッピング理論等)では、新生児は生得的に親の発話の韻律的特徴に敏感であるということが主張されている。よって、実験の被験者のような一般人についても、発話の韻律情報に敏感であるということは、自然な結果であるといえよう。

4.2 失敗例との比較

前章において、教示理解を獲得した組の教示理解プロセスを詳しく観察した。教示理解を成功した組と失敗した組とでは、どのような点が異なっているのだろうか？

まず、いかなる教示も理解できなかった組(HitRatio:0.29)にもかかわらず教示を無視するようにラケットを動かすと方向理解を獲得できた組(HitRatio:0.51)とを比較する。図8の太線は、各10試行毎の教示の種類を表したものである。このグラフは、成功した組において教示種類が徐々に減っており(図8の左)、失敗した組

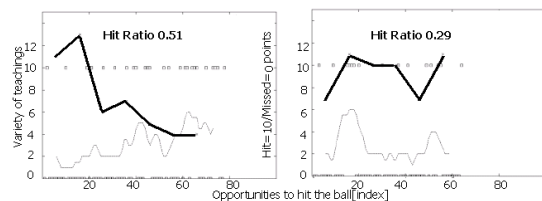


図8 Variety of instructions in successful/unsuccessful pairs

においてはそうはなっていないことを示している(図8の右)。しかしながら、これらの組の被験者において、典型的な教示の種類である「上上...」「下下...」発話のピッチ曲線と実際に教示音声を見た場合の聞き方において大きな差はみられなかった。成功した組においては、教示者が教示の種類を「上上...」「下下...」の二種類に減らしたことで、操作者はラケットのアクション二種類(右 or 左)とこの二種類の教示との「二対二」の対応付けを探索することで方向教示理解を得ることができる。しかし、失敗した組においては、教示種類は依然多いままであるので、多種類の教示中に成功した組が使用していた「上上...」「下下...」教示が存在したとしても、「二対二」の関係よりもさらに広い探索空間を持つ「多対二」の対応付けを模索する必要がある。結果として、この組の操作者は成功した組が意味理解に用いた教示に注目して自身の行動と対応づけることができなかった。ここで、方向理解を獲得するために重要な条件の一つとして、教示者が教示の種類を減らすことの重要性に気が付いているかどうか、ということが挙げられた。

この方向理解の達成は、教示者側の学習のみに依存しているのではなく、操作者の学習も重要であり、実際に成功・失敗の組の間において操作者の行動にも違いが観察された。成功した組の被験者はゲームの開始直後は積極的にラケットを動かさず、教示者の教示を推測することに集中し、教示の意味の予測をある程度した後に、ラケットを積極的に動かし始めていた。つまり、この操作者の動作は与えられた教示に対する自身の理解度を示していた。失敗した組の被験者においても、実験の開始時は成功した組の場合と同様、操作者がいきなり大きく動作することはなかったが、時間が経つにつれてこの操作者は教示が与えられていない

また、方向理解を獲得した9組中、4組は距離理解を獲得することができなかった。このフェーズにおい

でも、成功・失敗の被験者組の間に行動の違いが観察された。例えば、成功した組の教示者は、繰り返し発話で教示を与えることが操作者の教示理解において重要であるということを理解し、実際にこのことを操作者の反応から獲得していた。しかしながら、失敗した組の操作者はこの重要性に気かずに、むしろ「なぜ単発で教示を与えたときには逆の方向に動くのだろう?」と、ただ疑問に思うだけであった。

成功した組と失敗した組との行動比較から、このゲーム実験において高得点をとる条件、すなわちスムーズなコミュニケーションを成立する条件が明らかになった。つまり、インタラクションをする二者は相手の行動に対して適切に反応することを学習しなければならないということである。成功している組に見られるこれら経時的な行動の変化は、相互適応の結果であったと言える。

5. おわりに

本研究におけるゲーム実験環境において被験者間にスムーズなコミュニケーションを成立するために、(1)、教示者は、自身の与えた教示の結果としての操作者のラケット動作に十分な注意を払う必要があり、その観察の結果からその操作者に対して適切な教示を与える必要がある、(2)、操作者は与えられた教示に対してラケットを動かすことで、その教示に対する理解度を教示者に示す必要がある、ということが確認された。つまり、インタラクションをする二者がお互いに相手に対して適切な反応を返すことを学習する必要があった。ここでの、教示者と操作者の一連の行動の変化は、相互適応のプロセスと捕らえることができる。

教示音声における韻律的特徴に関して、方向理解を獲得しようとしているフェーズにおいては、高ピッチ情報は否定の意味として機能し、距離理解フェーズにおいては急かしの意味として機能している。ここで、高ピッチ要素は、操作者に対して一貫してその行動に注意を促す役割をしているということが言えた。そして、その具体的な意味が、操作者の教示理解のレベルによって「否定」と理解されたり「促進」と理解されたりしているという興味深い現象が観察された。

この研究において、「相互適応と発話中の韻律情報がコミュニケーションを構築するのに重要な役割を果たしている」という筆者らの仮説が実証された。この結果は、あらかじめ言語体系のルールの一部を埋め込むことで人とのコミュニケーションを構成するような従来の人-機械間インターフェイス思想に対して一石を投じることになるであろう。

謝辞 実験で使用したピッチ抽出プログラム開発において、松下技研株式会社の小沼氏に多大なご協力を頂いた。さらに、ゲーム実験環境においてはカリフォルニア大学サンディエゴ校の Agarwal 氏、松下電器産業株式会社の森川氏が開発したゲームソフト MATRIX を使用させて頂いた。この場を借りて皆様に感謝の意を表す。

参 考 文 献

- 1) D'Odorico, L.(1984). Non-segmental features in prelinguistic communication: An analysis of some types of infant cry and non-cry vocalizations, *Child Language*, 11, 17-27.
- 2) Fernald, A., & McRoberts, G.(1996). Prosodic bootstrapping: A critical analysis of the argument and the evidence. In J.L.Morgan and K. Demuth(Eds), *Signal to syntax: Bootstrapping from speech to syntax in early acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 3) Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I.(1989). A cross language study of prosodic modifications in mother's and father's speech to preverbal infants. *Child Language*, 16, 477-501